# DOCUMENT RESUME

| | |
|---|---|
| ED 443 286 | FL 026 324 |

| | |
|---|---|
| AUTHOR | Lindberg, Carl-Erik, Ed.; Lund, Steffen Nordahl, Ed. |
| TITLE | Scandinavian Conference of Linguistics, Volumes I [and] II (17th, Nyborg, Denmark, August 20-22, 1998). Odense Working Papers in Language and Communication. |
| INSTITUTION | Odense Univ. (Denmark). Inst. of Language and Communication. |
| REPORT NO | OWPLC-19 |
| PUB DATE | 2000-04-00 |
| NOTE | 460p. |
| AVAILABLE FROM | Institute of Language and Communication, University of Southern Denmark, Main Campus: Odense University, Campusveg 55, DK-5230 Odense M., Denmark. |
| PUB TYPE | Collected Works - Proceedings (021) |
| LANGUAGE | English, German, Danish |
| EDRS PRICE | MF01/PC19 Plus Postage. |
| DESCRIPTORS | *Applied Linguistics; Cognitive Processes; Communicative Competence (Languages); Computer Uses in Education; Danish; English (Second Language); Estonian; Finnish; Foreign Countries; German; *Grammar; Norwegian; Oral Language; Portuguese; Second Language Instruction; Second Language Learning; Sociolinguistics; Swedish; Syntax; *Uncommonly Taught Languages |
| IDENTIFIERS | Denmark; Estonia; Finland; Germany; Grice (H P); Norway; Portugal; Sweden |

ABSTRACT

The two volumes of these working papers include articles by linguists from Scandinavia and other parts of the world. Under the heading, "The Use of IT in Grammatical Analysis/Parsing" are five articles: "Creating Inflecting Electronic Thesauri"; "Tagging Speech Data--Constraint Grammar Analysis of Spoken Portuguese"; "A Constraint-Based Tagger for Norwegian"; "Swedish Constraint Grammar"; and "Two Two-Level Models of Norwegian." Under the heading "Functional/Cognitive Linguistics" are 12 articles: "The Subject is Nominative! On Obsolete Axioms and Their Deep Rootedness"; "Negation: A Cognitive Hurdle in Interpreting"; "Inflection Classes, Norwegian Verbs, and Cognitive Grammar"; "Conditionals in Institutional Settings"; "Different Competencies? A Comparison of the Productive and Receptive Languages of a Danish Agrammatic Patient"; "Some Reflections on Syntactic Structure and the Acquisition Process"; "Placement and Scope of Mainland Scandinavian Modal Adverbs"; "On Grammaticalization of Body-Part Nouns: 'Hand' in Finnish and Estonian"; "Some Pragmatic Properties of French Cleft Constructions"; "The Construal of the Scene: Selected Existential Constructions"; and one article each in German and Danish. Under the heading "Language in Interaction" are eight articles: "Overlap Management and Interactional Competence"; "Stereotypes in Action"; "Danish in Greenland -- a new Dialect? Preliminary Results From a Pilot Project"; "Meanings of the Word 'Work' in a Danish Unemployment Office"; "Codeswitching between Persian and Swedish"; and three articles in Danish. Under the heading "Endangered languages" are four articles: "How Fast Will a Language Die When It Is Officially No Longer Spoken?"; "Komi, an Endangered Language?"; "Phonetic Salience and Language Change: Verb Paradigms in Inari Sami"; and "Linguistic and Political Aspects of Alphabet-Making for a Threatened Language." All articles include a brief abstract in English. References appear throughout the articles. (KFT)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 443 286

# 17th Scandinavian Conference of Linguistics, Vol. 1[and] 2.

Odense Working Papers in Language and
Communication, No. 19, April 2000.

Carl-Erik Lindberg and Steffen Nordahl Lund, Eds.

University of Southern Denmark.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL HAS
BEEN GRANTED BY

Johannes
Wagner

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

FL 026 324

UNIVERSITY OF SOUTHERN DENMARK

# 17th Scandinavian Conference of Linguistics, vol. I

Carl-Erik Lindberg
& Steffen Nordahl Lund (eds.)

No. 19
April
2000

ODENSE
WORKING PAPERS
IN LANGUAGE
AND COMMUNICATION

Carl-Erik Lindberg & Steffen Nordahl Lund (eds.)
17th Scandinavian Conference of Linguistics, vol. I

Odense Working Papers in Language and Communication
No 19, vol I, April 2000

General editor: Steffen Nordahl Lund (snl@language.sdu.dk)

Institute of Language and Communication
University of Southern Denmark,
Main Campus: Odense University
Campusvej 55
DK-5230 Odense M.
(lfn@language.sdu.dk)

Printed in Denmark by Odense Universitets Trykkeri

# Indhold

Forord i

## The Use of IT in Grammatical Analysis/Parsing

*Antti Arppe, Mari Voipio, Malene Würtz*
Creating inflecting electronic thesauri 1

*Eckhard Bick*
"Tagging Speech Data" –
Constraint Grammar Analysis of Spoken Portuguese 11

*Kristin Hagen, Janne Bondi Johannessen, Anders Nøklestad*
A Constraint-Based Tagger for Norwegian 31

*Juhani Birn*
Swedish Constraint Grammar 49

*Sjur Nørstebø Moshagen*
BmNorTWOL and NyNorTWOL –
Two Two-level Models of Norwegian 75

## Functional/Cognitive Linguistics

*Jóhanna Barðal*
The Subject is Nominative!
On Obsolete Axioms and their Deep-Rootedness 93

*Anne Marie Bülow-Møller*
Negation: a cognitive hurdle in interpreting 119

*Hans Olav Enger*
Inflection classes, Norwegian verbs and Cognitive Grammar 127

*Tarja Salmi-Tolonen*
Conditionals in Institutional Settings 139

*Elisabeth Engberg-Pedersen, Lise Randrup Jensen*
Different competences? A comparison of the productive and receptive languages of a Danish agrammatic patient 153

*Volkmar Engerer*
Was läßt sich im Deutschen *aufhören* und *beenden*? Aspekt, Aktionsart und nominale Komplementation intervallabschlußdenotierender Phasenverben 171

*Peter Harder*
Some reflections on syntactic structure and the acquisition process 187

*Henrik Jørgensen*
Placement and Scope of Mainland Scandinavian Modal Adverbs 203

*Oili Karihalme*
Om begreppet Kognitiv situation 223

*Krista Ojutkangas*
On grammaticalization of body-part nouns: *käsi* 'hand' in Finnish and Estonian 237

*Eija Suomela-Salmi*
Some pragmatic properties of French cleft constructions 249

*Piotr Twardzisz*
The construal of the scene: Selected existential constructions 265

## Forord

Institut for Sprog og Kommunikation, Odense Universitet, arrangerede i dagene 20.-22. august 1998 det 17. Nordiske Lingvistmøde på Hotel Nyborg Strand ved Storebælt.

Denne række lingvistmøder arrangeres skiftevis af de enkelte nordiske lande med ca. 1 1/2 års mellemrum, og har gennem årene opnået nærmest institutionslignende status. Møderne er åbne for foredrag inden for samtlige lingvistiske discipliner; dog har arrangørerne mulighed for at udvælge og prioritere et antal hovedtemaer. Til lingvistmødet 1998 havde de lokale arrangører prioriteret følgende emner:

- The Use of IT in Grammatical Analysis/Parsing
- Functional/Cognitive Linguistics
- Language in Interaction
- Endangered Languages

Herudover var der en åben sektion med andre temaer.

I mødet deltog følgende hovedtalere, der alle er internationalt førende forskere inden for deres område: professor Östen Dahl, Stockholms Universitet, Professor John Taylor, University of Otago, New Zealand, professor Fred Karlsson, Helsingfors Universitet, professor Jürgen Streeck, University of Texas at Austin og professor Erling Wande, Stockholms Universitet.

I dette års lingvistmøde deltog ca. 80 sprogforskere, først og fremmest fra Skandinavien, men også fra det øvrige Europa, fra USA og New Zealand.

Det var på forslag fra præsidenten for the Nordic Association of Linguists, Helge Dyvik, Universitetet i Bergen, at området Truede Sprog blev optaget i dette års program. Området omfatter mindre sprog og især sprog som af historiske, sociale eller politiske grunde er ved at ophøre med at eksistere. I den "sprogudjævning" som i disse år foregår over hele verden, ikke alene ved udligning mellem rigssprog og dialekter, men også ved bl. a. "majorisering", hvor enkelte mindre sprog eller minoritetssprog trues af tilsigtet eller utilsigtet udryddelse, er en lingvistik og tillige kulturvidenskabelig beskæftigelse med dette område væsentlig, såvel ud fra et sprogligt og kulturelt, som ud fra et samfundsmæssigt og demokratisk synspunkt.

De fleste af deltagerne i 1998 årsmødet bidrog med foredrag over emner inden for de fire hovedtemaer. En række af disse foredrag udgives herved i serien *Odense Working Papers in Language and Communication* (hovedtalernes bidrag vil blive publiceret andetsteds).

# THE USE OF IT IN GRAMMATICAL ANALYSIS/PARSING

# Creating inflecting electronic thesauri

*Antti Arppe, Mari Voipio, Malene Würtz*

### Abstract

The contents of Swedish, Danish and Norwegian Bokmål electronic dictionaries were integrated with computerized morphological models for each respective language in order to create inflecting thesauri. Studying and analyzing these outcomes provided practical insight into the workings of lexical generality and synonymy. As a result, it would seem that lexical generality does not apply as universally as one would expect: in addition to categorical, morphology-based restrictions and exceptions, there are also pragmatic, common-sense restrictions to it. Furthermore, it would seem that synonymy does not categorically extend and survive to all the respective inflected forms of the original synonyms, i.e. that inflectional features of words should also be taken into account when describing synonymy.

## 1. Introduction

The work described in this article was carried out by the authors at Lingsoft, Inc., a language engineering company based in Helsinki, Finland. Lingsoft specializes in providing high-quality natural language processing tools for word processing, text retrieval and information management systems.

One group of such tools is inflecting electronic thesauri, which Lingsoft has developed for the major Nordic languages (in order of the time of completion: Finnish, Swedish, Danish and Norwegian Bokmål). In this presentation we will start with a general overview of what inflecting thesauri are and how they were developed by us. Then we will concentrate on linguistic problems we encountered with the inflecting electronic thesauri for Danish, Norwegian Bokmål and Swedish.

## 2. What we did

Electronic thesauri are certainly not our invention: any dictionary profits from the search capabilities that a computerized form can provide. Several text processing programs have long had built-in thesauri, but until now they have only worked with the base forms, i.e. they have been facsimiles of the original paper versions of the dictionaries. Lingsoft's novelty is to provide the possible synonyms of a word with the same inflection as the input word, which is the definition of an inflecting thesaurus in a nutshell. This feature makes these tools easier to use while writing or editing any normal running text.

Schematically, an inflecting thesaurus works in the following way:

I. inflected word or phrase & morphological analysis
 ->base form(s) of the word
 ->inflectional information of the word

II. base form(s) & data structure containing synonym relationships
 -> list of synonyms

III. lists of synonyms & inflectional information & morphological generation
 -> list of synonyms in inflected form

A concrete example of this is the following synonym list given for klassificerades, either a past passive or a genitive form of the past participle of klassificera 'to classify':

klassificerades
-> ordnades (verb)
 delades upp i klasser, indelades, klassades, grupperades, rangordnades, rankades, inrangerades, graderades, kategoriserades, systematiserades

-> betecknades (verb)
bedömdes
-> betecknades (adjectival participle)
bedömdas

The basic ingredients of an inflecting thesaurus is a computational morphological model of a language – in our case the two-level models or TWOLs (Koskenniemi 1983) – and electronic versions of synonym dictionaries – typically licensed from well-known book publishers. The Danish two-level lexicon had been licensed earlier from Thomas Bilgram and developed further by Lingsoft (Bilgram 1994), while the Norwegian Bokmål and Swedish lexicons were developed entirely at Lingsoft (Moshagen 1998) and (Karlsson 1992). The synonym lists were licensed from Politikens Forlag for Danish (Politikens Synonymordbog 'PSO'), Cappelens Forlag for Norwegian Bokmål (Cappelens Media Synonymordbok) and Walters Lexikon for Swedish (published as Bonniers Synonymordbok).

Developing the inflecting electronic thesauri was a process that lasted several months, partly because of the unexpected linguistic problems that had to be solved in order to obtain a satisfactory result. Lingsoft's inflecting thesauri for Danish and Norwegian Bokmål were finished in May 1997 and they are included in the Danish and Norwegian versions of the first upgrade version of Danish and Norwegian Office 97, called Office 97A. Lingsoft's inflecting thesaurus for Swedish was finished already in 1996, and was included in the Swedish version of Office 97.

## 3. What we found out

### 3.1 Concept of lexical generality

The theory of generative morphology postulates that it is 'in general possible to attach to any word the entire set of inflections associated with the word [base form] in question' (Scalise 1984). In other words, a set of inflectional categories can be generalized and extended to hold for all – both existing and prospective – members of a particular lexical category or word class. Bybee (1985) uses the term *lexical generality* for this phenomenon.

For example, all Swedish nouns can be inflected in principle in eight different forms (case, definiteness and number). Thus, were we to form an entirely novel Swedish noun such as rucka, its inflectional paradigm would consequently be rucka, ruckan, ruckor, ruckorna, ruckas, ruckans, ruckors, ruckornas, whatever this word may eventually mean.

However, lexical generality does not apply across all derivations. Thus, from the Swedish verb tävla 'to compete' we get tävling 'competition', but from finna 'to find' we do not get *finning but finnande 'finding'. Even less so can all words be freely compounded with others.

Any rule-based morphological model of a language, such as a two-level model, can be seen as an incarnation of lexical generality, since such a model incorporates a lexicon of word stems, each of which is assigned to a particular lexical category, and sets of inflectional rules, i.e. paradigms, deemed to apply for all the members of a particular lexical category.

### 3.2 Extent of lexical generality

Thus, an inflecting thesaurus can be seen as the application of lexical generality over the entire contents of the synonym dictionary. Since both computerized morphological models *AND* electronic versions of synonym dictionaries of the respective languages have co-existed only recently, the effects of this combination on a large scale have not been practically observable until now, which we can see with inflecting thesauri.

During the development of the inflecting thesauri, however, we found that lexical generality does not always hold. There are certain limits to it, both categorical and pragmatic. To create truly functional, i.e. user-friendly, inflecting thesauri we therefore found it necessary to implement certain grammatical restrictions in the productive inflectional mechanism.

Table 1

| | *Test 1* | *Test 15* |
|---|---|---|
| Total number of suggestions | 2823 | 1893 |
| Average number of suggestions per word | 36 | 24 |
| Words with no suggestions | 13<br>(16.6 %) | 0 |
| Wrong word forms suggested | 76<br>(2.7 %) | 3<br>(0.16 %) |
| Odd word forms suggested | 126<br>(4.5 %) | 7<br>(0.48 %) |

To illustrate the necessity of these restrictions, we present in Table 1 some figures derived from the Danish test list, containing 78 randomly selected words from PSO, used to detect mistakes and other irregularities. The suggested synonyms generated in the first test contained quite a few wrong word forms and even more odd suggestions, and one-sixth of the words in the test list received no synonyms at all. The wrong word forms were mostly due to errors in DANTWOL's inflectional system and were easily corrected or eliminated. The odd suggestions, however, were mostly comparative or superlative forms of adjectives and passive forms of multiword phrases. We chose to eliminate only the ones that sounded very strange to the native ear, but the validity of the rest of these suggestions could certainly be discussed as well.

Categorical limits:

Some verbs simply cannot appear in the past passive: *vars, *blevs, *gikks or in the imperative: *aftn!, *bero!, as these forms are semantically impossible. As for comparison of adjectives, it cannot be applied to so-called compound adjectives such as kœmpestor, skidegod or non-inflecting adjectives such as ekstra.

Pragmatic limits:

Past passive forms of verbs like binde and lade (that is, bandtes and lodes) are morphologically correct, but sound rather strange to the native user; they are beyond a pragmatic limit. Likewise, imperative forms of the verbs figurere and eksistere seem odd. Norwegians also find some morphologically and semantically correct imperative forms like sykl! quite strange. In this category we also placed the genitive forms of adjectives in the singular – stors and storts are in principle correct forms, but they are pragmatically dubious, as are the genitives of past participle singular indefinite forms (spists, elskets). Another example of pragmatic limits is the formation of comparative and superlative forms of certain adjectives - some adjectives describing colours can easily be compared (eg. grønnere), while others are doubtful, to say the least (eg. *orangere).

### *3.3 Inflectional homonymy*

During the testing process, we also became aware of the *inflectional homonymy* which is rather abundant in the Scandinavian languages. The testing was carried out by running lists of inflected test words gathered randomly from the original thesaurus sources through the initial electronic versions of the thesauri to ensure that these really were fully inflecting. The output of these tests was then carefully scrutinised in order to detect as many irregularities as possible. An example of inflectional homonymy is a word like flyver that can be a verb in the present tense as well as a noun in base form. Likewise, the word anklagedes can be interpreted as either the past passive of the verb anklage or as the definite genitive form of the participle anklaget (or the definite/indefinite plural form...).

Inflectional homonymy is not necessarily a problem; for example flyver as a noun and flyver as the present tense of the verb flyve are both common and well-known forms. But when some of the inflectional homonyms are common forms and others are very uncommon or unlikely, they look confusing to the user of the thesauri. For example, by typing gås the user most likely means the bird goose and not the present passive form of the verb gå 'to go'. This is one of the reasons why pragmatic limits HAVE to be drawn somewhere.

### *3.4 Synonymy and inflection*

Another phenomenon that was observed in the creation of these inflecting thesauri was how well synonymy extends with inflection. Traditional theory on synonymy focuses on the semantic traits of words and the grade of synonymy in a semantic sense (absolute versus partial synonymy), but does not really address questions regarding inflection. For instance:

> '... two items are synonymous, if they have the same sense.' (Lyons 1968)
>
> 'Synonyms, then, are lexical items whose senses are identical in respect of "central" semantic traits, but differ, if at all, only in what we may provisionally describe as "minor" or "peripheral" traits.' (Cruse 1986)

> 'Synonymy occurs when two terms correspond in all three aspects of *designatum*, *connotation* and *range of application*. [...] If the correspondence is exact in one or two of the three aspects but not in all, the two words are near synonyms.' (Zgusta 1971)

Existing synonym dictionaries are by and large composed to represent semantic relationships of the exact word forms that they contain. For most cases, synonymy appeared to extend quite well into the inflected forms of synonym words. For instance, let us look at the [near] synonymous nouns bok 'book' and verk 'printed work', and the near synonymous adjectives gammal 'old' and möglig 'moldy' in the following sentences

Denna *bok är gammal.* 'This book is old'
-> *Detta verk är mögligt.* 'This printed work is moldy'
*De här böckerna är äldre.* 'These books are older'
-> *De här verken är mögligare.* 'These printed works are moldier'

In the above cases, synonymy can be seen to extend to the inflected forms of different declinations of nouns, with bok belonging to the utrum gender and verk belong to the neuter gender.

On the other hand, quite a few cases of verb synonymy are sensitive to inflection. Take for example hitta 'to find' and finna 'to find', and säga 'to say' and betyda 'to mean'. The following sentences are more or less synonymous:

Sen hittade jag boken. 'Then I found the book.'
-> Sen fann jag boken. 'Then I found the book.'
Vad vill det säga? 'What does it mean? [Literally: What does it say?]'
-> Vad vill det betyda? 'What does it mean?'

In the following sentences, on the other hand, the inflected forms of finna and säga are no longer appropriate synonyms for various reasons, though they are morphologically interpretable, if not altogether acceptable:

Boken hittades. 'The book was found.'
–> Boken fanns. 'The book existed.'
Säg någonting! 'Say something!'
–> *Betyd någonting! 'Mean something!'

In the first case, inflection of finna into the passive form changes the pragmatic meaning from 'find' to 'exist' or 'there is'. In the second case, inflection of betyda into the imperative form sounds odd to the native Swedish ear.

In the Swedish inflecting thesaurus, this degeneration or discontinuity of synonymy as a result of inflection was observed particularly with the passive and imperative forms of verbs, which therefore had to be manually reviewed. Similar phenomena were also observed with the Danish and the Norwegian Bokmål thesauri. All in all, it was our conclusion that synonymy cannot be categorically expected to extend to all the inflected forms of words originally deemed synonyms, which probably is not a big surprise to most lexicographers. If one would want to present reliable synonyms for passive forms of verbs, for instance, one would have to compose an entirely new synonym dictionary for that purpose, using the passive forms of verbs as the starting point.

## 4. Conclusion

Creating inflecting thesauri for Danish, Norwegian Bokmål and Swedish by combining computerized morphological models with electronic versions of synonym dictionaries gave ample practical insight into a few linguistic phenomena, namely lexical morphology and lexical semantics (synonymy) at their intersection. The most essential implications that we discovered in this work are the following:

- lexical generality does not apply as universally as one would expect: in addition to categorical, morphology-based restrictions and exceptions there are also pragmatic, common-sense restrictions to it.
- synonymy does not categorically extend to all the respective inflected forms of the original synonyms.

*Lingsoft, Inc.*
*Tehtaankatu 27-29 K, FIN-00150 Helsinki, Finland*
*antti.arppe@lingsoft,mari.voipio@helsinki.fi, malene.wuertz@helsinki.fi*

## References

Bilgram, Thomas. 1994. Computerstyret analyse af dansk. En praktisk analyse af en væsentlig kilde til homonymi i dansk med forslag til kontekstuel disambiguering ved hjælp af constraint-regler. Upubliceret speciale, Institut for Lingvistik, Aarhus Universitet.
Bybee, Joan L. 1985. Morphology: John Benjamins Publishing Company.
Cruse, D A. 1986. Lexical Semantics. Cambridge: CUP.
Karlsson, Fred. 1992. SWETWOL: A Comprehensive Morphological Analyzer for Swedish. Nordic Journal of Linguistics 15, 1992. 1-45.
Koskenniemi, Kimmo. 1983. Two-Level Morphology: A General Computational Model for Word-Form Recognition and Production. In: Publications of the Department of General Linguistics 11. Helsinki: Department of General Linguistics, University of Helsinki.
Lyons, John. 1968. Introduction to Theoretical Linguistics. Cambridge: CUP.
Moshagen, Sjur. Morphological Analyzers for Norwegian (Bokmål and Nynorsk): Paper presented at the 17th Scandinavian Conference on Linguistics, August 20-22, 1998.
Scalise, Sergio. 1984. Generative Morphology. The Netherlands: Foris Publications.

Zgusta, Ladislav. 1971. Manual of lexicography. Janua Linguarum, Series Maior: 39. The Hague: Mouton.

# "Tagging Speech Data" –

# Constraint Grammar Analysis of Spoken Portuguese

*Eckhard Bick*

**Abstract**
The paper discusses the automatic grammatical analysis of spoken language data for Portuguese. A Constraint Grammar based tagger/parser for written Portuguese (Bick 1996 and 1997) was used as a point of departure and run on transcribed portions of a Brazilian urban speech corpus, NURC (*"Norma Lingüística Urbana Culta,* e.g. Castilho et. al., 1989 and 1993). Quantitative evaluation of tagging results showed a stable performance (error rates under 1%) for both written and speech data, while the 2-3% syntactic error rate of the original text parser deteriorated considerably when the same rules were applied to preprocessed speech data (8-9% error rate). However, by introducing additional rules and by disambiguating pauses (in-utterance) and breaks (inter-utterance), error rates could be brought down to 4-5%, suggesting the applicability of the CG appoach to (transcribed) speech data.

## 1. Introduction

The paper raises the question whether automatic grammatical parsers designed for written language input can be made to handle transcribed spoken language data. A Constraint Grammar based tagger/parser for written Portuguese (Bick 1996 and 1997) was used as a point of departure and run on transcribed portions of a Brazilian urban speech corpus, NURC (*"Norma Lingüística Urbana Culta,* e.g. Castilho et. al., 1989 and 1993). By designing a transcription specific preprocessor as well as by adding new rules and modifying old ones, the original system was to be adapted so as to handle both spoken and written language input.

Constraint Grammar (CG) as both a disambiguation based parsing technique and a notationally flat dependency grammar was introduced by Fred Karlsson (Karlsson et al. 1995) and has been applied to a variety of languages, including - besides English - Swedish (Birn 1998), Norwegian (Hagen et. al. 1998) and Portuguese (Bick 1996). CG rules are disambiguation rules and syntactic mapping rules that are applied to morphologically analysed (but still ambiguous) text, differing from probabilistic Hidden Markov Model analysers both in their primarily linguistic nature and their wide scope of context conditions, the context window usually being an entire sentence, not only immediately neighbouring words. A typical (simplified) rule for English would for example discard a direct object reading in favour of a subject reading, if there is no transitive verb to the left, and the next word to the right is a finite verb:

REMOVE (@OBJ) (0 @SUBJ) (1 C VFIN) (NOT *-1 <vt>)

It is this kind of wide context rule one would expect - when moving from written language to speech data input - to be affected by the speech specific lack of punctuation and clear sentence boundaries.

## 2. The point of departure

The original Portuguese parser yields correctness rates (defined as recall percentages at near 100% disambiguation) - on unrestricted unknown text - of over 99% for morphology/PoS and 97-98% for syntax (Bick, 1997:2), a performance which makes the system suitable for applications like unsupervised parsing, corpus searches, interactive grammar teaching and - experimental - MT (cf. test site at http://visl.hum.ou.dk/), which have so far all been aimed exclusively at written language. While the morphological/PoS tagger module proved quite robust in test runs on spoken language data (with a success rate of around 99% even without additional rules), syntactic analysis fared somewhat worse, with an initial correctness rate of 91-92% for the - rule-wise - unmodified system.

In order to explain this discrepancy between morphological robustness and syntactic failure, hypotheses like the following can be formulated and subsequently tested by changing the system's preprocessor and rule system accordingly:

- In my parser, rules with **morphological targets** mostly use a **shorter context range** (group structure) than those with syntactic targets, cf. table (rule scope). Thus the proportion of rules without and with unbounded contexts is 10 times as high for rules targeting morphological tags as for syntactic targets, and 70-80% of all syntactic rules stretch their context all the way to the sentence delimiters – making these rules vulnerable to the speech specific absence or vagueness of such delimiters.

**Table: rule scope**

| | morphological targets | | | | syntactic targets | | | | all |
|---|---|---|---|---|---|---|---|---|---|
| | safe | heuristic levels | | | safe | heuristic levels | | | |
| | | 1. | 2. | 3. | | 1. | 2. | 3. | |
| **REMOVE** tag<br>(only local contexts) | 403 | 112 | 13 | 27 | 153 | 37 | 4 | 2 | 651 |
| **REMOVE** tag<br>(≥ 1 global contexts) | 183 | 44 | 5 | 5 | 941 | 219 | 17 | 1 | 1415 |
| *local/global* | **2.2** | 2.5 | 2.6 | 5.4 | **0.2** | 0.2 | 0.2 | 2.0 | **0.5** |
| **SELECT** tag<br>(only local contexts) | 271 | 70 | 8 | 7 | 60 | 2 | 1 | 1 | 420 |
| **SELECT** tag<br>(≥ 1 global contexts) | 129 | 23 | 9 | 2 | 209 | 57 | 3 | - | 432 |
| *local/global* | **2.1** | 3.0 | 0.9 | 3.5 | **0.3** | 0.0 | 0.3 | - | **1.0** |

- **Incomplete utterances** tend to leave group structure intact more often than clause structure, - at least if one doesn't count repetitional modifications/corrections of prenominal modifiers *(essas esses progressos, esta este caminho, da dos nomes)*, where word-class adjacency rules can often override agreement rules.

- Speech data **lacks punctuation** and has **unclear sentence window** borders, which is especially bad for syntactic CG analysis, which tends to use many unbounded context restrictions (cp. 1).

- Speech data is filled with "**syntactic noise**", repetitions and false starts of one- or two-word chunks, as well as pause and phatic interjections *(ahn, uh, eh, etc.)*.

## 3. Preprocessor tasks

In order to address these problems and improve syntactic performance, a preprocessor was designed with the specific goal of establishing utterance or sentence boundary candidates and removing syntactic noise.

**1. Orthography and layout normalisation** (character set, line numbers)

**2. Repetitions and false starts** (automatically commented out by $-signs)

*mas é vo/ voluntária né? -->mas é <$vo/> voluntária né?*

*então então vem tudo aquilo de cambulhada e im/ e im/ im::POSto sobre nós*

*--> então <$então> vem tudo aquilo de cambulhada e <$im/> <$e> <$im/> <stress> imposto sobre nós*

*com o com o nome da pesso/ da do do escritor lá*

*--> com o <$com> <$o> nome da <$pesso/> <$da> do <$do> escritor lá*

**3. Phonetics**

- Vowel length markers are removed, e.g. *u::ma pessoa --> uma pessoa*

- In-word stress marking is commented out, e.g. *esnoBAR --> <stress> esnobar*

**4. Introducing "dishesion marker candidates"** (ccc)

- Due to a complete lack of full stops, colons and commas (only question marks [?] and turn taking [¶] are used), other means of marking syntactic windows become necessary, and strings like ..., eh, éh, () are marked as "dishesion elements", as well as quotes if they enclose more than 1 word.

### 5. Mapping "dishesion markers" as

**a) <break>** (major syntactic break, clause or sentence boundary)

* <break> markers can be used by the CG rules to establish maximal group size or valency scope; e.g., <break> should not occur between a premodifier and its head, or between main verb and direct object.

**b) <pause>** (non-word hesitation/pause marker)

* <pause> markers are not allowed to break up group or clause continuity.

### 6. Disambiguating "dishesion markers" (eee)

***a) "xxx" --> eee xxx eee --> <pause> xxx <pause>***

If a single word is surrounded by dishesion markers, these are treated as <pause>

***b) eee (e) que/quando/embora ... --> <pause>***

If a dishesion marker is followed by a conjunction or relative, possibly with an interfering coordinator, it is treated as <pause>.

***c) que/quando/embora ... eee --> <pause>***

If a dishesion marker is preceded by a conjunction or relative, it is treated as <pause>

***d) eee + PRP --> <pause>***

If a dishesion marker is followed by certain prepositions (de, em, com, sem, por), it is to be treated as <pause>

***e) PRP/det + eee + NON-art/dem --> <pause>***

If a dishesion marker is preceded by a preposition or a determiner (or a fused preposition+determiner), it is to be treated as <pause>, unless it is directly followed by an article or demonstrative (in which case the <pause>/<break> ambiguity is retained)

## 4. Grammar tasks

The next effort concerned the CG rule grammar as such, where dishesion marker candidates had to be integrated in those tag sets that denote possible syntactic breaking points. The PAUSE set, for example, includes not only the dishesion marker, but also certain interjections:

```
LIST PAUSE = "uhn" "ahn" "eh" "eee" <pause> <break> IN ;
```

The <break> tag is useful in NON-sets since these are often used in BARRIER conditions in CG-rules, baring group attachment, for instance:

```
LIST NON-NP = PERS SPEC ADV VFIN INF PRP KS KC <rel>
<interr> "<$\,>" <break> >>> <<< ;
```

On the sentence level, <break> is a potential clause boundary marker, the same way certain complementizers, comma and hyphen are:

```
LIST CLB = KS <interr> <rel> "<$\,>" "<$->" KOMMA <break> ;
```

Next, rules have to be crafted for the disambiguation of cohesion markers - are they breaks denoting "sentence" window borders, or just pauses embedded in the syntactic flow of speech?

For instance, dishesion markers are not <break> (but <pause>) if they intervene:

(a) between a "name bearer" and its name: *o rei $$ Alfonso*

(b) between a noun and the preposition 'de': *pai $$ de muitos filhos*

(c) between an intensifier and an attribute: *uma maneira um pouco $$ calcada*

(d) between a noun and a potential postmodifier or object complement of the same gender and number: *estou vendo* ***a TV*** *evidentemente $$ muito* ***presa*** *a ...*

(e) between a transitive main verb and its direct object.

```
(a) REMOVE (<break>) (-1 (<+n>)) (1 <*>)

(b) REMOVE (<break>) (-1 N) (1 PRP-DE)

(c) REMOVE (<break>) (-1 <quant>) (1 ATTR/<attr>)

(d) REMOVE (<break>) (*-1 NFP BARRIER ALLuPAUSE/ADV) (*1 ATTR-FP
    BARRIER ALLuPAUSE/ADV)

(e) REMOVE (<break>) (-1C @MV LINK 0 <vt>) (*1C @<ACC BARRIER @NON-
    >N)
```

Of course, the use of dishesion markers and their introduction in NON-sets and CLB-sets, has to be balanced between the advantages of providing better defined windows of analysis, and the draw-backs of disallowing many long range rule contexts that are conditioned by CLB-barriers and the like.

## 5. Special problems

### a) Premodifier clashes (da dos)

In a simple correctional article clash ('comeu a o bolo') both articles will receive the @>N (premodifier) tag, but in more complex cases there may be problems, for instance, where a preposition is repeated as well. Here, the first determiner will be analysed as @P< (argument of preposition).

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| eu | não | estou | agora | por | dentro | **de** | **a** | **de** | **os** | nomes | sabe | ? |
| SUBJ> | ADVL> | FMV | ADVL> | <SC | P< | A< | **P<** | N< | >N | P< | FMV | |

**b) "Faulty" noun phrases: stranded premodifiers in incomplete np's** (um, uma) **and agreement errors** (codificação nada normativo)

* In speech more than in text, a long distance between head and modifier may result in agreement lapses (here, M - F).

* stranded premodifiers tend to assume np-head function in a syntactic parse, which may seem odd, but is hard to avoid, and may well be the logical solution - after all, in a word-based tagger/parser there are no zero constituents, and every function has to be attached *somewhere*.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| e | $e | não | havendo | uma | codificação | não | $pause |
| CO | | ADVL> | IMV | >N | <ACC | ADVL> | |
| | | | ICL-ADVL> | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| $break | $eee | **um** | **uma** | $pause | nada | **normativo** |
| | | <ACC | <ACC | | >A | N< |

Agreement failure (here SG - PL) does occur in adjacent position, too. The examples are taken from a transscription where the speaker (a lecturer) admitted to being nervous on being taped.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| a | demanda | de | moeda | por | transação | $pause | é | $pause |
| >N | SUBJ> | N< | P< | N< | P< | | FMV | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **principal** | **motivo** | por | **os=quais** | as | pessoas | $pause | retêm | moeda |
| >N | <SC | ADVL> | P< | >N | SUBJ> | | FMV | <ACC |
| | | | FS-N< | | | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| nós | podemos | resumir | isso | **em** | **um** | **exemplinhos** | numérico |
| | | | | <ADVL | >N | P< | N< |

### c) True ambiguity with regard to primary sentence constituents

| estão | gravando | agora | este | $pause | está | passando | $está passando |
|---|---|---|---|---|---|---|---|
| FAUX | IMV | <ADVL | <ACC |  | FAUX | IMV |  |
|  | ICL-AUX< |  |  |  |  | ICL<AUX |  |

| agora | em | São=Paulo | O | Grito | não | é | ? |
|---|---|---|---|---|---|---|---|
| <ADVL | <ADVL | P< | >N | <SUBJ | ADVL> | FMV |  |
|  |  |  |  | <ACC |  |  |  |
|  |  |  |  | <SC |  |  |  |

Note that the stranded premodifier 'este' receives the function tag of a presumed np-head.

### d) Difficulties in identifying subjects:

Consider the following example, where three subject tags have to be found and tolerated in the same speech chunk without clear clause boundaries: *televisão, ela, telespectador*:

| porque | a | **televisão** | sendo | estatal | **ela** | é | muito | $stress |
|---|---|---|---|---|---|---|---|---|
| SUB | >N | SUBJ> | IMV | <SC | SUBJ> | FMV | >A |  |
| FS-<ADVL |  |  | ICL-<ADVL |  |  |  |  |  |

| uniformizada | $pause | $break | não | há | espectáculos | diversificados | o |
|---|---|---|---|---|---|---|---|
| <SC |  |  | ADVL> | FMV | <ACC | N< | >N |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **telespectador** | $pause | $break | o | fica | sempre | $pause | preso |
| SUBJ><br><ACC | | | ACC> | FMV | <ADVL | | <SC |

| | | | | | |
|---|---|---|---|---|---|
| a | filmes | ou | a | $a | conferências |
| A<PIV | P< | CO | <PIV | | P< |

Here, '**ela**' is semantically anaphoric to 'televisão', which syntactically belongs to its own non-finite subclause. '**telespectador**' lacks a sentence/analysis window marker (before its article), which is why function has not been fully disambiguated in this case. '**o**' before the main verb 'fica' might be part of yet another subject candidate with only its article left, but since the grammar strongly disallows adjacency of articles and finite verbs, 'o' is treated as a personal pronoun in the accusative. 'o' does not bear any meaning in this sentence, and would be ignored by a human listener, but once uttered and transcribed, the word has to be handled in the grammar one way or another.

**e) Notationally caused and interaction based errors in multi-speaker data**

In a notation that uses only one time line, utterances of speaker S2 may syntactically "cut" an utterance of speaker S1. Also, speakers S1 and S2 may interact syntactically, finishing each other's groups or clauses. In the example, 'adequado' (S1) is subject complement (SC) for 'está' (S2), 'perfeitamente' (S2) is premodifier (>N) for 'adequado' (S1):

*L2 para aquele ... está* ***perfeitamente*** *...*

*L1* ***adequado***

*L2* ***adeQUAdo:: do*** *... é muito mais interessante ... é uma*

*[*

*L1()*

*L2 grande oportunidade para os nossos artistas não é ?*

*L1 isso é muito bom:: eh:: e ain/ e:: e a novela puxa o disco porque parece que na vendagem dos discos eles são muito ... requisitados esses discos de novelas né ?*

*L2 H. você escreveu qualquer coisa muito interessante*

### 6. Positive side effects: robustness

CG's flat analysis is very robust, and especially advantageous with unclear sentence boundaries or nested sentences, both of which are frequent in speech data. Consider the 5 main verbs in the following comma- and coordinator-free sentence:

e **é** uma grande atriz $break então **choca** demais $pause $break

CO **FMV** >N >N <SC ADVL> **FMV** <ACC

aquela paulista $stress quatrocentona **que** ele **faz** bem $stress

>N <SUBJ N< ACC> SUBJ> **FMV** >A

FS-N<

grifado $break aliás de uma maneira um=pouco $pause calcada

<OC ADVL> ADVL> >N P< >A N<

demais **porque** esse tipo **acho** **que** já se **diluiu**

A< SUB >N SUBJ> **FMV** SUB ADVL> ACC> **FMV**

FS-<ACC

nem **existe** mais $pause mas ...

<ADVL **FMV** <ADVL CO

Even double main verbs without any sensible syntactic analysis, and breeches of the uniqueness principle are tolerated fairly well by the CG-grammar:

$break isto é **levava** a um tipo de vida nômade

SUBJ> **FMV** **FMV** <PIV >N P< N< P< N<

When all goes well, the system tolerates overlapping clauses with double uncoordinated subjects and a shared direct object, as well as - to a certain degree - complex and interrupted np's and np-modifiers (boxes).

| | **problems:** | |
|---|---|---|
| papai | N M S @SUBJ> | |
| mesmo | DET M S @N< | |
| tem | V PR 3S IND @FMVobligatorily transitive verb *without* direct object | |
| em | PRP @<ADVL | |
| os | DET M P @>N | |
| <$nos> | | |
| livros | N M P @P< | |
| de | PRP @N< | |
| ele | PERS M 3S NOM/PIV @P< | |
| ele | PERS M 3S NOM/PIV @SUBJ> 2 subjects without co- or subordination | |
| tem | V PR 3S IND @FMV | 2 main verbs without co- or subordination |
| muitas | DET F P @>N | |
| expressões | N F P @<ACC | direct object serving verbs in 2 clauses |
| $pause | | |
| completamente | ADV @>A | |
| caidas | V PCP F P @N<heavy postnominal with adjunct and argument | |

| | | |
|---|---|---|
| em=desuso | VPP @A<PIV | |
| e | KC @CO | |
| portuguesas | N F P @<ACC?? | less heavy postnominal *after* heavy postnominal |
| e | KC @CO | |
| <$por/> | | |
| e | KC @CO | |
| $pause | | |
| de | PRP @SC> @N< | very distant pp-postnominal with false start |
| português | N M S @P< | |
| clássico | ADJ M S @N< | |

| | | |
|---|---|---|
| não | ADV @ADVL> | |
| é | V PR 3S IND @FMV | finite clause without punctuation or header |
| $? | | |

## 7. Performance

A quantitative comparison of the two versions of the parser (written language vs. speech data) yielded the following results, with correctness defined as *recall at near 100% disambiguation*, counting both false tags, missing tags and false ambiguity as errors.

**Parser performance on running text** (VEJA news magazine and fiction)

| **Text:** | *O tesouro*<br>ca. 2500 words | | *VEJA 1*<br>ca. 4800 words | | *VEJA 2*<br>ca. 3140 words | |
|---|---|---|---|---|---|---|
| **Error types:** | errors | correct | errors | correct | errors | correct |
| Part-of-speech errors | 16 | | 15 | | 24 | |
| Base-form & flexion errors | 1 | | 2 | | 2 | |
| **All morphological errors** | 17 | **99.3 %** | 17 | **99.7 %** | 26 | **99.2 %** |
| syntactic: word & phrases | 54 | | 118 | | 101 | |
| syntactic: subclauses | 10 | | 11 | | 13 | |
| **All syntactic errors** | 64 | **97.4 %** | 129 | **97.3 %** | 114 | **96.4 %** |
| "local" syntactic errors due to PoS/morphological errors | - 27 | | - 23 | | - 28 | |
| **Purely syntactic errors** | 37 | **98.5 %** | 106 | **97.8 %** | 86 | **97.3 %** |

**Parser performance on speech data (before/after grammar adaptation)**

(NURC *[norma lingüística urbana culta]*, São Paulo)

| **speech sample** | **sample size** | **morphological correctness** | **syntactic correctness** |
|---|---|---|---|
| 2 speaker dialogue<br>(topic: cinema, television, actors)<br>females, 60 yrs (journalist and writer) | 2810 words | 99.2 % | 95.7 % |
| secondary school teaching monologue (history), female 36 yrs | 2080 words | 99.5 % | 96.3 % |
| university teaching monologue (economics), male 31 yrs | 1600 words | 99.0 % | 95.4 % |
| *base line:*<br>2 speaker dialogue (same as above) analysed with unmodified grammar | 1100 words | 98.9 % | 92.6% |

## 8. Conclusion

While an automatic parser originally designed for written Portuguese was able to more or less maintain its performance on speech data *morphology* (word class, etc.), error rates tripled for speech data *syntax*. Judging from the effectiveness of according rule changes and preprocessing, one can conclude that at least one of the reasons for this striking difference resides in the fact that the disambiguation of morphological ambiguity involves mostly short range group context that is left intact even in the grammatically often incomplete utterances of spoken language, while rule based syntactic analysis depends on long range context patterns, working less than perfectly without a clear sentence window, without full complementation of obligatory valency, and with breaches of the uniqueness principle. The hypothesis was tested by tagging - through a preprocessor module - what I call *dishesion markers* ("...", "eh", etc.) in the corpus as both <pause> and <break> for later disambiguation, thus introducing "sentence boundary" candidates, which may be disambiguated by either crude word form context or elaborate long range CG rules. Once disambiguated, the <break> markers provide more "traditional" syntactic window delimiters for the system's Constraint Grammar, considerably improving syntactic tag recall. Examples where modification of the syntactic rules as such proved necessary are violations of the uniqueness principle due to iterations or modified ("corrected") iterations, or cases, where one speaker complements the valency pattern of a syntactic unit uttered by another speaker. Especially problematic are clashes, where a speaker strands dependents without their heads (for instance, subjects without a verb, or a premodifier without its nominal head) and departs on a new syntactic path.

Preliminary quantitative results suggest that break markers and rule modifications can narrow the gap between the parser's performance on written and spoken Portuguese, respectively, to a few percentage points for syntax and nearly eliminate it for part of speech tagging.

## References

Bick, Eckhard, 1996. "Automatic Parsing of Portuguese", in *Proceedings of the Second Workshop on Computational Processing of Written Portuguese*. Curitiba:CEFET-PR.

Bick, Eckhard, 1997a. "Dependensstrukturer i Constraint Grammar Syntaks for Portugisisk", in: Brøndsted, Tom & Lytje, Inger (eds), *Sprog og Multimedier*, Aalborg: Aalborg Universitetsforlag.

Bick, Eckhard, 1997b. "Automatisk analyse af portugisisk skriftsprog", in: Jensen, Per Anker & Jørgensen, Stig. W. & Hørning, Anette (eds), *Danske ph.d.-projekter i datalingvistik, formel lingvistik og sprogteknologi*, pp. 22-20, Kolding:Institut for erhvervssprog og sporglig informatik, Handelshøjskole Syd.

Birn, Juhani, 1998. "Swedish Constraint Grammar", in *Proceedings of the 17th Scandinavian Conference on Linguistics*, this volume. Odense.

Castilho, Ataliba Teixeira de (ed.), 1989. *Português culto falado no Brasil*. Campinas: Editora da Unicamp.

Castilho, Ataliba de (ed.), 1993. *Gramática do Português Falado*, vol. 3, Campinas: Editora da Unicamp.

Hagen, Kristin and Janne Bondi and Anders Nøklestad, 1998. "A Constraint-Based Tagger for Norwegian", in *Proceedings of the 17th Scandinavian Conference on Linguistics*, this volume. Odense.

Karlsson, Fred & Voutilainen, Atro & Heikkilä, Juka & Anttila, Arto (eds.), 1995. *Constraint Grammar, A Language-Independent System for Parsing Unrestricted Text*. Berlin: Mouton de Gruyter.

# A Constraint-Based Tagger for Norwegian

*Kristin Hagen, Janne Bondi Johannessen and Anders Nøklestad*

**Abstract**

Disambiguating morphosyntactic taggers are computer programs which provide the words in a text with grammatical information and which are able to pick the correct reading for ambiguous words based on linguistic context. We describe such a tagger for Norwegian BOKMÅL and NYNORSK which is based on the Constraint Grammar formalism (Karlsson et al. 1995). The tagger disambiguates through the use of linguistic constraints that operate only the level of individual words, which means that no phrase structure is established. We show how it is possible to perform morphological and syntactic disambiguation of Norwegian texts without having recourse to a phrasal level.

## 1. Introduction

Automatic taggers can greatly enhance the value of text corpora in linguistic research by providing morphosyntactic information about the words in the text, thus allowing users to include specifications of grammatical categories in their text search. At the University of Oslo, work is currently being carried out to develop an automatic morphosyntactic tagger for Norwegian BOKMÅL and NYNORSK. The project comprises seven man-labour years and extends from April 1996 to December 1998.[1]

## 2. Morphosyntactic taggers

### 2.1 General

A morphosyntactic tagger is a computer program whose task is to provide correct grammatical information for every word in any stretch of running text. Ideally, every word should have a grammatical description, and the description should be unambiguous. Thus, the tagger must be able to disambiguate words that have more than one reading, like *så* in (1):

(1) *På trikken til byen i dag så jeg henne igjen*
'On tram-the to town-the today saw I her again'
*så*: SE verb past tense/ SÅ conj./ SÅ adv./ SÅ noun masc. appellative sg. indef./
SÅ verb imperative/ SÅ verb infinitive

The tagger should have the ability to decide that *så* is a past tense verb in this context.

Our disambiguating tagger consists of the following parts:

(2) I. Pre-processor

- Splits the text into sentences
  - uses special lists of abbreviations to decide when a full stop is a period and when it belongs to an abbreviation
  - decides when a word sequence is a headline rather than a sentence, based, among other things, on the number of words in the sequence

II. Multi-tagger

- Uses a full form list, i.e. a list of inflected words, to mark each word form with all tags that are appropriate for it
- Puts tags on remaining delimiters like commas, colons and dashes
- Analyses unknown words in a compounding module, using linguistic rules and lists of full forms and stems

- Marks unknown capitalised words as proper nouns

III. Morphological disambiguator

IV. Morphosyntactic mapping

- Performs mappings from morphological description to a set of possible syntactic functions for each word in the text

V. Syntactic disambiguator

In this paper we will concentrate on parts III to V, i.e. morphological disambiguation, morphosyntactic mapping and syntactic disambiguation.

For these parts we are using a formalism called Constraint Grammar, which has been developed at the University of Helsinki by Fred Karlsson and co-workers (Karlsson et al. 1995). The computer implementation of Constraint Grammar that we are using has been developed by the Finnish company Lingsoft. Constraint Grammar is similar to so-called Dependency grammar. A special feature of this kind of grammar is that it does not split larger grammatical structures into hierarchical sub-structures. Phrases like NP or VP do not occur in this grammar; only relationships between individual words are described.

## 2.2 Statistical goals

If the tagger provides the correct tags for each word, it reaches what is called one hundred percent RECALL. Furthermore, if no words receive any incorrect tags, we obtain one hundred percent PRECISION. Although one hundred percent recall and precision are the ideal goals for the tagger, we are not aware of any other taggers that have reached such high numbers. Thus, a more realistic goal might be about 96% precision and a recall close to 100%.

However, the larger the number of morphosyntactic distinctions that is reflected in the set of tags, the more difficult it will be to reach a high degree of disambiguation. We have nevertheless chosen a considerably larger set of tags than what seems to have been used with comparable taggers for other languages, because we want to retain as many distinctions as possible in the tagger's output. At the moment the tagger has reached a recall of 99.2% and a precision of 94.8% for BOKMÅL, and 98.9% recall and 92.5% precision for NYNORSK, but these numbers should increase somewhat before the end of the project.

## 3. Morphological disambiguation - how to find the finite verb

We will now discuss morphological disambiguation of Norwegian texts, focusing on the particular problem of finding finite verbs (we will see later on that this is important for subsequent syntactic disambiguation). In Norwegian the finite verb functions as the second constituent in declarative sentences. As first constituent we can find all kinds of phrases containing an indefinite number of words and phrases:

(3) Different NPs as first constituent:

a. *Du går ...*

'You walk ...'

b. *En sliten mann går ...*

'A tired man walks ...'

c. *Mannen går* ...

'The man walks ...'

d. *Mannen i parken går* ...

'The man in the park walks ...'

e. *Mannen med brun frakk og svart hatt går* ...

'The man with a brown coat and a black hat walks ...'

f. *Mannen med en frakk som er så skitten at den stinker, går* ...

'The man with a coat that is so dirty that it stinks walks ...'

g. *Mannen med den evnen at han kan snakke med månen, går* ...

'The man with the skill that he can talk to the moon walks ...'

(4) Different AdvPs as first constituent:

a. *Da går du* ...

'Then you walk ...'

b. *I dag og i morgen går du* ...

'Today and tomorrow you walk ...'

c. *Ofte går du* ...

'Often you walk ...'

d. *Heller går du* ...

'Rather you walk (= you rather walk) ...'

e. *Enten går du* ...

'Either you walk ...'

f. *Ganske ofte går du* ...

'Quite often you walk ...'

Since we are using the Constraint Grammar model without a phrasal level, it is difficult to determine where the first constituent ends. Many words are also multiply ambiguous with both verbal and non-verbal readings, as shown in (5), which is an example of a multi-tagged text that has not yet been disambiguated.

(5) Ambiguous example: adverb as first constituent:

*Likevel koster huset mange penger*

'Still costs house-the many money-PL'

'Still the house costs a lot of money'

| | |
|---|---|
| *Likevel* | LIKEVEL adv |
| *koster* | KOST subst mask appell fl ub/KOSTE verb pres |
| *huset* | HUS subst nøyt appell ent be/ |
| | HUSE adj <perf-part> mask fem ub ent/ |

| | |
|---|---|
| | HUSE adj <perf-part> nøyt ub ent/HUSE verb perf-part/ |
| | HUSE verb pret |
| *mange* | MANGE adj pos ub be fl |
| *penger* | PENGE subst mask appell fl ub |

Here, both *koster* and *huset* could be the finite verb of the sentence.

Since the constraints must be formulated locally, without reference to phrases or other hierarchical structures, we have developed a large number of detailed constraints for each possible correct reading. Today we have about 90 constraints for verbs in the present or past tense. One of them, shown in (6), deals with sentences like the one above.[11]

(6) Constraint: adverb as first constituent:

Choose the 'verb pres'-reading if you can find nothing but adverbs between a clause boundary and the potential verb.

This constraint is used by the tagger to choose *koster* as the finite verb of the sentence, as we see in (7):

| | |
|---|---|
| (7) *Likevel* | LIKEVEL adv |
| *koster* | KOSTE verb pres |
| *huset* | HUS subst nøyt appell ent be |
| *mange* | MANGE adj pos ub be fl |
| *penger* | PENGE subst mask appell fl ub |

We have tried to make the constraints as specific as possible in order to avoid mistaggings. For instance, we have a constraint for sentences where the first constituent consists of an NP containing a PP, like the one shown in (8). A constraint of this kind must of course be more complicated, as we see in (9). (10) shows the result of using this constraint.

(8) Ambiguous example: NP with PP as first constituent:

*Det gamle huset med den store verandaen kostet mange penger*

'The old house-the with the big veranda-the cost many money-PL.'

| | |
|---|---|
| *Det* | DET det dem nøyt ent/DET pron pers 3 ent nøyt |
| *gamle* | GAMMAL adj pos mask fem nøyt be ent/ |
| | GAMMAL adj pos ub be fl/GAMMEL adj pos ub be fl/ |
| | GAMMEL adj pos mask fem nøyt be ent |
| *huset* | HUS subst nøyt appell ent be/HUSE verb pret/ |
| | HUSE adj <perf-part> mask fem ub ent/ |
| | HUSE adj <perf-part> nøyt ub ent/HUSE verb perf-part/ |
| *med* | MED prep |
| *den* | DEN det dem fem ent/DEN det dem mask ent/ |

| | |
|---|---|
| | DEN pron pers 3 ent mask fem |
| *store* | STOR adj pos mask fem nøyt be ent/ STOR adj pos ub be fl |
| *verandaen* | VERANDA subst mask appell ent be |
| *kostet* | KOST subst nøyt appell ent be/KOSTE verb perf-part/<br>KOSTE verb pret/KOSTE adj <perf-part> mask fem ub ent/<br>KOSTE adj <perf-part> nøyt ub ent |
| *mange* | MANGE adj pos ub be fl |
| *penger* | PENGE subst mask appell fl ub |

(9) Constraint: NP with PP as first constituent:

Choose the 'verb pret' reading if the word is preceded by a clause boundary followed by a sequence of adverbs, adjectives, nouns, pronouns, conjunctions, determiners and/or prepositions, and the first word in the sequence is not a preposition.

(10)Disambiguated example: NP with PP as first constituent:

*Det gamle huset med den store verandaen kostet mange penger*

'The old house-the with the big veranda-the cost many money-PL'

| | |
|---|---|
| *Det* | DET det dem nøyt ent |
| *gamle* | GAMMAL adj pos mask fem nøyt be ent/ |
| | GAMMEL adj pos mask fem nøyt be ent |
| *huset* | HUS subst nøyt appell ent be |
| *med* | MED prep |
| *den* | DEN det dem mask ent |
| *store* | STOR adj pos mask fem nøyt be ent |
| *verandaen* | VERANDA subst mask appell ent be |
| *kostet* | KOSTE verb pret |
| *mange* | MANGE adj pos ub be fl |
| *penger* | PENGE subst mask appell fl ub |

## 4. Syntactic disambiguation

As we saw in (2), the last two components of the system perform syntactic analysis of the text. Thus, while morphological constraints are used for deciding on a morphological description, the last components try to decide what syntactic function these words have in the sentence. Syntactic analysis requires that morphological disambiguation is carried out first, since it would be impossible to determine syntactic functions like Subject and Object without knowing, for instance, which word is the finite verb of the sentence.

Like the morphological constraints, the syntax used in Constraint Grammar is word based. This means that no hierarchical phrase structure is constructed. Instead, phrasal heads are marked with syntactic main functions like Subject or Object, while other words in the phrase are simply marked as modifiers of some word to the left or right in the phrase. This is illustrated in (11)[iii]. The Subject of the sentence is the noun phrase *det gamle huset* 'the old house', but only the head of the phrase, i.e. *huset*, is marked as Subject. The other two words in the phrase, *det* and *gamle*, are marked as modifying determiner and adjective, respectively, with arrows pointing in the direction of the head they modify – that is, *huset*.

(11) 'The old house cost a lot of money'

| | | | | | |
|---|---|---|---|---|---|
| *Det* | *gamle* | *huset* | *kostet* | *mange* | *penger* |
| @DET> | @ADJ> | @SUBJ | @FV | @ADJ> | @OBJ |

As was shown in (2), syntactic analysis is done in two stages. In the first stage, the morphological descriptions for each word are mapped to a set of possible syntactic tags. Most mapping rules put some constraints on the word's environment, like the one shown in (12), which was used for marking *gamle* 'old' in (11) as a modifying adjective:

(12) Morphosyntactic mapping for definite neuter singular adjectives:

If a definite neuter singular adjective is directly followed by a word that can be a neuter singular noun, and that cannot be anything but a noun, mark the adjective as a modifier of the noun.

As (12) illustrates, the mapping rules can specify context in terms of morphological descriptions. This is sufficient for giving unambiguous syntactic labels to verbs and to most modifiers. Thus, after morphosyntactic mapping has been carried out, the words in our example sentence will have the syntactic labels shown in (13), where we have included a PP as part of the Subject NP:

(13) *Det gamle huset med den store verandaen kostet mange penger*

'The old house with the big veranda cost a lot of money'

| | |
|---|---|
| *Det* | @DET> |
| *gamle* | @ADJ> |
| *huset* | @SUBJ @OBJ @I-OBJ @S-PRED @<P-UTFYLL @TITTEL @SUBST> @APP @LØS-NP |
| *med* | @ADV |
| *den* | @DET> |
| *store* | @ADJ> |
| *verandaen* | @SUBJ @OBJ @I-OBJ @S-PRED @<P-UTFYLL @TITTEL @SUBST> @APP @LØS-NP |
| *kostet* | @FV |
| *mange* | @ADJ> |
| *penger* | @SUBJ @OBJ @I-OBJ @S-PRED @<P-UTFYLL @TITTEL @SUBST> @APP @LØS-NP |

In order to get any further with the analysis, however, we must be able to refer to other syntactic labels in the sentence. The last syntactic component, the SYNTACTIC DISAMBIGUATOR, enables us to do this. This component contains constraints that choose syntactic labels by referring to other syntactic labels and/or the morphological tags of other words in the sentence. We will see later on that the position of words relative to the finite verb of the sentence can be used to determine functions like Subject and Object. First, however, we need to make sure that only one word in each phrase is available as the carrier of a syntactic main function. Otherwise we could, for instance, be running the risk of choosing *verandaen* ('the veranda') as Subject in our example sentence, since this is the word that occurs immediately in front of the finite verb. Thus, we first run a number of syntactic constraints that will check whether any of the words are modifiers. One of these is shown in (14); this constraint will identify *verandaen* as the complement of the preposition *med* 'with' in our example sentence.

(14) Syntactic constraint for complement of preposition:

If there is an unambiguous preposition to the left and there are no syntactic labels other than pre-modifiers (@DET>, @ADJ>, @ADV>, @SUBST>, @TITTEL and @KON) or Adverbials between the preposition and the word we are examining, mark the word as a complement of the preposition.

Note that this constraint refers both to purely morphological information, such as the part of speech PREPOSITION, and to syntactic labels like pre-modifying determiner and Adverbial.

When this constraint has been applied to the sentence, we get the set of labels shown in (15):

| | | |
|---|---|---|
| (15) | *Det* | @DET> |
| | *gamle* | @ADJ> |
| | *huset* | @SUBJ @OBJ @I-OBJ @S-PRED @<P-UTFYLL @TITTEL @SUBST> @APP @LØS-NP |

| | |
|---|---|
| *med* | @ADV |
| *den* | @DET> |
| *store* | @ADJ> |
| *verandaen* | @<P-UTFYLL |
| *kostet* | @FV |
| *mange* | @ADJ> |
| *penger* | @SUBJ @OBJ @I-OBJ @S-PRED @<P-UTFYLL @TITTEL @SUBST> @APP @LØS-NP |

What remains now is to determine the syntactic functions of *huset* and *penger*. The constraint that chooses Subject function for *huset* is based on the fact that indefinite nouns occurring after the verb rarely function as Subject, so that in such cases the Subject is more likely to be found in front of the verb. The constraint is fairly complicated, but what it does is essentially this:

(16) Syntactic constraint for Subject:

For a noun, nominative pronoun or determiner that occurs before the finite verb, choose the Subject label if there is an indefinite common noun after the verb as long as that noun can have some syntactic function other than Subject.

We see that in this constraint the position of the word we are examining, as well as the position of the other, indefinite, noun, are defined with respect to the finite verb of the sentence. This shows how important it is for the syntactic analysis that the verb was identified during morphological disambiguation.

Another point worth noting about this constraint is that it is HEURISTIC. This means that we cannot be one hundred percent certain that the word will in fact be a Subject in this context, and so this constraint will only be applied if no other, more reliable constraint has already chosen a Subject for the sentence. (In fact, by the time [16] is applied, a number of other constraints will already have removed some of the other possible labels listed for *huset* in [13] and [15]).

Having decided on the Subject of the sentence, we are in a position to identify the syntactic function of the word *penger*, through the use of the following constraint:

(17) Syntactic constraint for Direct Object:

For a word that occurs after the finite verb, choose the Direct Object label if there is an unambiguous Subject in front of the finite verb, there are no copula verbs in the sentence, and the word we are examining cannot be an Adverbial.

Again, we see that the constraint uses the position of the finite verb as the point of reference for the position of the word we are examining, as well as for the position of the Subject.

Thus, when the Subject and Direct Object constraints have been applied, we arrive at the set of labels shown in (18), which constitutes the final analysis of the sentence:

| | | |
|---|---|---|
| (18) | *Det* | @DET> |
| | *gamle* | @ADJ> |
| | *huset* | @SUBJ |

| | |
|---|---|
| *med* | @ADV |
| *den* | @DET> |
| *store* | @ADJ> |
| *verandaen* | @<P-UTFYLL |
| *kostet* | @FV |
| *mange* | @ADJ> |
| *penger* | @OBJ |

## 5. Conclusion

In this paper we have shown how morphological and syntactic disambiguation of Norwegian texts is feasible even without recourse to a phrasal level. Such disambiguation is important because it makes it possible to tag large amounts of text which can be used, for instance, in linguistic research. Since no tagged Norwegian corpus is currently available (with the exception of a hand-tagged corpus of Ibsen texts) we think that the development of the tagger we have presented here will make a valuable contribution to Norwegian language research.

*The Text Laboratory, University of Oslo*
*Department of Linguistics, P.O. Box 1102 Blindern*
*0317 Oslo, Norway*
*e-mail: kristin.hagen@inl.uio.no, j.m.b.johannessen@ilf.uio.no,*
*anders.noklestad@ilf.uio.no*

## Reference

Karlsson, Fred, Voutilainen, Atro, Heikkilä, Juha, and Anttila, Arto (ed.) 1995. *Constraint Grammar: A Language-Independent System for Parsing Unrestricted Text*. Berlin: Mouton de Gruyter.

---

[i] The project is mainly financed by the Norwegian Research Council, the Documentation Project, the Text Laboratory and the Finnish company Lingsoft. Our lexicons are based on *Bokmålsordboka* and *Nynorskordboka* as well as on electronic lexicons provided by IBM.

[ii] In this paper we will concentrate on informal explanations of what the constraints do, rather than on the actual form of the constraints.

[iii] Explanations for syntatic tags used in the text: @ADJ>: pre-modifying adjective; @ADV: Adverbial or post-modifying prepositional phrase; @ADV>: pre-modifying adverb; @APP: apposition; @DET>: pre-modifying determiner; @FV: finite verb; @I-OBJ: Indirect Object; @KON: conjunction; @LØS-NP: NP without syntactic function; @OBJ: Direct Object; @<P-UTFYLL: complement of preposition; @S-PRED: Subject Complement; @SUBJ: Subject; @SUBST>: pre-modifying noun; @TITTEL: title.

# Swedish Constraint Grammar

*Juhani Birn*

## Abstract

Swedish Constraint Grammar (SWECG) is a system for part-of-speech disambiguation and shallow syntactic analysis of running Swedish text, developed within the Constraint Grammar (CG) frame-work proposed by Fred Karlsson (1990). The basic principles and the formalism of CG, as applied especially to the description of English, are presented in full in Karlsson et al. (1995). The lexical basis of SWECG is defined by the morphological analyser SWETWOL, designed according to the principles of Kimmo Koskenniemi's (1983) two-level model. The basic properties of SWETWOL are presented in Karlsson (1992).

After an introductory note on the short "history" of context-sensitive disambiguators of Swedish, this presentation consists of two main sections. Section 2, System overview, outlines the different components of SWECG and, along the way, reports on their present status. Section 3, Lexicon for disambiguation and syntax, is concerned with fundamental grammatical considerations within the SWECG framework. In particular, the treatment of determiners and participles is reviewed. Grammatical tags used in SWECG are listed in an Appendix.

## Introduction

In his book *Svensk lexikografi*, published in 1994, Sven-Göran Malmgren comments on the state of computational analysers of Swedish. To illustrate, Malmgren (1994: 121) uses the following text:

> Nu skärper kommunerna kraven på föräldrar som inte betalar daghemsavgifter. Ända sedan 70-talet har kommunerna haft rätt att stänga av barn om avgiften uteblivit – men först nu har man börjat göra det. Orsaken till den hårdnande attityden är kommunernas försämrade ekonomi.

Many of the words in the above text are lexically ambiguous (or, to use Malmgren's term, they are homographs). E.g. *rätt* can be a noun, an adjective, or an adverb, *man* can be a noun or a pronoun, and *försämrade* can be a verb or a participle. The point Malmgren makes is this: "Tack vare kon-texten väljer en mänsklig läsare av texten omedelbart den rätta av de teoretiskt möjliga analyserna. Men man har ännu inte lyckats överföra denna förmåga till ett analysprogram; även det bästa svenska analysprogrammet kan endast ge **alla** de möjliga analyserna av *rätt* etc. men inte välja ut den i sammanhanget riktiga [...]."

According to SWETWOL, the analyser that gives **all** the possible analyses, the 41 words of the above text have all in all 69 analyses. The text is fairly simple from a disambiguation point of view. Using contextual clues SWECG reduces the 69 analyses to 41, i.e. one for each word, and each of these analyses is the contextually correct one, e.g. *rätt* is analysed as a noun, *man* as a pronoun, and *försämrade* as a participle. (For lack of space the actual SWECG output is not given here.) So, what is going to be outlined below **is** a context-sensitive disambiguator of Swedish.

### System overview

The SWECG system consists of six sequentially applied components, summarised in (1-6) below. The first three (1-3) define the input to (SWE)CG proper, which is concerned with disambiguation and syntax (4-6). The goal of SWECG can be summarised thus: to pick for each word in the sentence, first, the contextually correct morphological reading (4) and, second, the syntactic label that describes the word's function in the containing clause (6).

Input: running text

**(1) Preprocessing**

Conversion of the text to the presupposed format

**(2) Morphological analysis**

SWETWOL analysis and recognition of morphological ambiguity (i.e. each word is assigned all of its possible readings)

**(3) Morhological heuristics**

Assignment of readings to words not recognised in (2)

**(4) Morphological disambiguation**

Elimination of contextually inappropriate readings due to (2) and (3)

**(5) Morphosyntactic mapping**

Recognition of syntactic ambiguity (i.e. each morphological reading is assigned one or more syntactic function labels)

**(6) Syntactic disambiguation**

Elimination of contextually inappropriate labels due to (5)

Output: analysed text

In automated part-of-speech disambiguation a distinction is often made between rule-based and stochastic (statistics-based) approaches. CG systems are of the rule-based type. The viability of the rule-based approach has been discussed extensively in the context of English CG, see e.g. Samu-elsson & Voutilainen (1997). For a stochastic tagger for Swedish see Carlberger & Kann (1998).

SWECG is now being developed at Lingsoft, Inc. Descriptive contributions to components (1-3) have been made by a number of people, both at the University of Helsinki and at Lingsoft (e.g. Juhani Birn, Fred Karlsson, Krister Lindén, Markku Norberg, Kari Pitkänen, Fredrik Westerlund). The original compilation of the SWETWOL lexicon was made by Fred Karlsson (1992). An early version of component (4), using the original SWETWOL, was made at the University of Helsinki. The current, Lingsoft-based descriptions pertaining to components (4-6) have been made by the present author over a period of seventeen months. As a preamble to that work, some SWECG-geared changes were made in the SWETWOL description (cf. section 3 below).

The tasks performed by the different components are illustrated below with a component-wise analysis of the following sentence, qualifying as input to the system in the form given.

Dessa entreprenöriella faktorer hade än så länge dämpat explosionen.

**(1) Preprocessing.** The preprocessor converts the text to the format presupposed by the lexical analyser. The tasks include the recognition of punctuation marks, abbreviations (e.g. *bl.a.*), and fixed syntagms (multi-word expressions treated as single words, e.g. *än så länge*). After pre-processing the example sentence looks as shown below. The graphic words that make up a fixed syntagm are glued together with the understroke '_'.

*dessa
entreprenöriella
faktorer
hade
än_så_länge
dämpat
explosionen
$.

According to the current listings, the preprocessor recognises some 150 fixed syntagns (far from what the number usefully could be) and more than 5000 abbreviations.

**(2) Morphological analysis.** The morphological analyser SWETWOL, using a lexicon which con-tains about 75,000 base-forms, provides all the possible readings for each word form it recognises. E.g. the word *dämpat*, as shown below, is two-way ambiguous: verb (V) or adjective (A). For *entre-prenöriella*, not recognised by SWETWOL, see component (3). For the tags see the Appendix.

```
"<*dessa>"
   "denna" <**c> <DEM> <MD> DET UTR/NEU DEF PL NOM @DN>
   "denna" <**c> <DEM> PRON UTR/NEU DEF PL NOM
"<entreprenöriella>"
"<faktorer>"
   "faktor" N UTR INDEF PL NOM
"<hade>"
   "ha" <AUX> V ACT PAST
"<än_så_länge>"
   "än_så_länge" <COLLOCATION> ADV
"<dämpat>"
   "dämpa" V ACT SUPINE
   "dämpa" <PCP2> A NEU INDEF SG NOM
"<explosionen>"
   "explosion" N UTR DEF SG NOM
"<$.>"
   "$." CLB <PUNCT>
```

A reading consists of a base-form, written in quotation marks (e.g. *"dämpa"*), and a list of features, given in the conventional order shown below (where '*' stands for zero or more). For SWECG, additional (angle-bracketed) features are relevant to the extent they are useful for disambiguation. Even syntactic function labels (prefixed by '@') may be assigned already in the lexicon.

< >* POS INFL* @FUNC*
where:
< >* = additional features
POS = part-of-speech, major syntactic category
INFL* = inflection/agreement categories, either nominal (gender,

definiteness, number, case) or verbal (voice, finiteness: tense or mood or nonfinite form)
@FUNC* = syntactic functions

The morphological ambiguity rate of Swedish as determined by SWETWOL defines the overall task facing the SWECG disambiguator. When Lingsoft's SWETWOL was applied to a newspaper corpus of some 1.9 million words (from *Svenska Dagbladet* 1996), it was found that 50.26% of the words were ambiguous (two or more readings), and that the average number of readings per word was 1.82. (These figures refer to the situation after so called local disambiguation of spurious compound and derivative readings; see Karlsson 1992. Prior to local disambiguation, the corre-sponding figures were 58.55% and 2.05.)

**(3) Morphological heuristics.** This is a module for the analysis of words not represented in the SWETWOL lexicon. The rules base their analyses on word shape. The word *entreprenöriella* is assumed to be an adjective on the basis of its ending, and it receives a SWETWOL-like adjective treatment. If no other rule applies, a noun reading is assigned as a default.

```
"<entreprenöriella>"
   "entreprenöriella" <NON-SWETWOL> A UTR/NEU DEF SG NOM
   "entreprenöriella" <NON-SWETWOL> A UTR/NEU DEF/INDEF PL NOM
```

The module contains at present only about 60 rules. It may be assumed, however, that incorrect part-of-speech assignments are rare, as most of the new words in a typical text are nouns (often names), analysed as such by the default rule. A test was made with a 10,600 word newspaper extract (*Svenska Dagbladet* 1996). Of these words, SWETWOL recognised 10,366 (97.8%), so 234 were left for morphological heuristics. Of these 234, according to the present author's judgement, 223 (95.3%) were nouns, and 11 non-nouns. The part-of-speech tag assigned by morphological heuristics was correct in 228 cases (97.4%), erroneous in 6 cases. All the errors were cases where a non-noun was analysed as a noun by the default rule.

**(4) Morphological disambiguation.** The task of the disambiguator is to discard contextually inappropriate readings. In the example sentence, the disambiguator's output is optimal: each word retains only one reading, and that reading is the contextually correct one.

```
"<*dessa>"
   "denna" <**c> <DEM> <MD> DET UTR/NEU DEF PL NOM @DN>
"<entreprenöriella>"
   "entreprenöriella" <NON-SWETWOL> A UTR/NEU DEF/INDEF PL NOM
"<faktorer>"
   "faktor" N UTR INDEF PL NOM
"<hade>"
   "ha" <AUX> V ACT PAST
"<än_så_länge>"
   "än_så_länge" <COLLOCATION> ADV
"<dämpat>"
   "dämpa" V ACT SUPINE
"<explosionen>"
   "explosion" N UTR DEF SG NOM
"<$.>"
   "$." CLB <PUNCT>
```

The rule formalism of CG will not be dealt with here. The following (somewhat simplified) prose paraphrase of one of the rules, viz. the one responsible for disambiguating *dämpat* as V in the example sentence, serves here to indicate the detailed, "atomistic" nature of many of the rules:

For a word, W, ambiguous between V SUPINE and <PCP2> A NEU, select the V reading
- if, somewhere to the left of W, there is a form of the verb "ha", and
- if, between "ha" and W, there is no other verb and no clause boundary indicator, and
- if W is not preceded by a neuter indefinite singular determiner or by a coordinator, and
- if W is not followed by a neuter indefinite singular noun or by a coordinator.

SUPINEs may of course occur also in other contexts (e.g. *bott* in *Jag känner alla som bott här*), and for these other contexts there are other rules. Obviously, it is not a goal of a CG to formalise any one rule or statement describing the overall distribution of a category (e.g. SUPINE in Swedish).

The current Swedish description contains about 2,100 morphological disambiguation rules. A constraint grammarian's task can be summarised thus: to try to optimise both the recall and the precision of the disambiguation rules. The notions recall and precision are here understood in the sense they are defined in Voutilainen (1995:172). **Recall**: the ratio 'received correct readings / all correct readings'. **Precision**: the ratio 'received correct readings / all received readings'. A recall of less than 100% indicates that some correct (i.e. intended) readings are lost, and a precision of less than 100% indicates that some superfluous readings remain in the output.

How do the SWECG morphological disambiguation rules perform in terms of recall and preci-sion? In the development of the rules we have been using a corpus of nearly 110,000 words (standard non-fiction). Here are, first, the performance rates maintained in this **development corpus**, given as percentages and as absolute figures. Recall: 99.86% (109788/109940). Precision: 96.72% (109788/114560). – These rates would indicate quite a satisfacory performance level, but they are not valid more generally, of course, because the corpus is the one used in the development of the rules.

SWECG has not been tested extensively for recall and precision with new texts. For this presen-tation, the following test was made. The test data consisted of four extracts **new to the system**, about 1000 words each. Two of the extracts come from the periodical *Folket i bild* (1995) (Fib.1 and Fib.2 below), and two from the daily *Svenska Dagbladet* (1996) (SvD.1 and SvD.2). First, in a SWETWOL analysis of the extracts, the readings considered to be the correct ones were marked manually by a linguist (the present author). Then, the SWECG program counted the recall and precision of the computational rules; the results are shown below. To these can be added that 96.59% (4304) of the words became fully disambiguated (one reading only).

| | Words | Recall | Precision | Readings/word |
|---|---|---|---|---|
| Fib.1 | 1016 | 99.61% (1012/1016) | 94.74% (1012/1065) | 1.048 |
| Fib.2 | 1253 | 99.60% (1248/1253) | 94.96% (1248/1309) | 1.045 |
| SvD.1 | 1075 | 99.81% (1073/1075) | 96.55% (1073/1101) | 1.024 |
| SvD.2 | 1112 | 99.28% (1104/1112) | 95.24% (1104/1156) | 1.040 |
| **Total** | 4456 | **99.57%** (4437/4456) | **95.36%** (4437/4631) | 1.040 |

The assessment of the above results is made problematic by a number of factors, among which are: the small size of the test data; the interpretation of what is 'correct' being dependent on the rule writer's judgement; and, if the results were to be usefully compared with those reported for other systems, the differences between the tag sets and, if that be the case, between the languages. Issues such as these we have not addressed.

In his dissertation Atro Voutilainen reported, on the basis of the analysis of large samples of text, that English CG had a recall / precision rate of 99.7–100% / 93–97% (Vouti-lainen 1995:186). Irrespective of how the performance figures of SWECG compare to those of other systems, it is clear to us that not everything that could be done for the development of SWECG has yet been done. In particular, a larger development corpus would contribute to a more accurate description.

**(5) Morphosyntactic mapping.** Next, morphological readings are provided with syntactic function labels. The labels, listed in the Appendix, are of three main types: (1) **verb function** labels (main verb or chain verb, finite or non-finte), (2) **head function** labels, assigned to the topmost head words of clause-level constituents (such as subject, object, adverbial), and (3) **modifier function** labels, either premodifiers (with the pointer '>' indicating that the head is found to the right) or postmodifiers (with the pointer '<' indicating that the head is found to the left). SWECG does not try to account for the function of clauses in sentences.

The basic idea here is to introduce ambiguity, meaning typically that a part-of-speech category is simply associated with all of its possible functions. For instance, the category A (adjective) is assumed to have the thirteen possible functions listed in the reading of *entreprenöriella* below. For the labels see the Appendix.

```
"<*dessa>"
   "denna" <**c> <DEM> <MD> DET UTR/NEU DEF PL NOM @DN>
"<entreprenöriella>"
   "entreprenöriella" <NON-SWETWOL> A UTR/NEU DEF/INDEF PL NOM @AN>
   @SCOMP @OCOMP
   @SUBJ @OBJ @IOBJ @AOBJ @ADVL @AD> @P>> @<P @<N @NPHR
"<faktorer>"
   "faktor" N UTR INDEF PL NOM @SUBJ @OBJ @IOBJ @AOBJ @SCOMP
   @OCOMP @ADVL @NN>@P>> @<P @<NN @NPHR
```

**(6) Syntactic disambiguation.** The last task is to discard contextually inappropriate function labels. In the example sentence, the system's output is optimal: each word retains only one label, and that label is the contextually correct one. The analysis reads thus: *Dessa* is a determiner (@DN>) and *entreprenöriella* is an adjective (@AN>) premodifying the subject head *faktorer* (@SUBJ), *hade* is a finite chain verb (@+FCV), *än_så_länge* is an adverbial (@ADVL), *dämpat* is a non-finite main verb (@+FMV), and *explosionen* is the object (@OBJ).

```
"<*dessa>"
    "denna" <**c> <DEM> <MD> DET UTR/NEU DEF PL NOM @DN>
"<entreprenöriella>"
    "entreprenöriella" <NON-SWETWOL> A UTR/NEU DEF/INDEF PL NOM @AN>

"<faktorer>"
    "faktor" N UTR INDEF PL NOM @SUBJ
"<hade>"
    "ha" <AUX> V ACT PAST @+FCV
"<än_så_länge>"
    "än_så_länge" <COLLOCATION> ADV @ADVL
"<dämpat>"
    "dämpa" V ACT SUPINE @-FMV
"<explosionen>"
    "explosion" N UTR DEF SG NOM @OBJ
"<$.>"
    "$." CLB <PUNCT> @
```

The SWECG syntactic disambiguation rules are not yet fully developed. In simple cases, e.g. the above example, the output is as intended, but in more complex cases much ambiguity remains unresolved. At present, the description contains about 400 rules; at least three or four times that amount would be needed before satisfactory results could be expected. It would be immature, at this stage of development, to report any performance figures.

In this section, the SWECG system has been presented in outline. In the next section, the focus is on fundamental grammatical considerations within the framework.

**Lexicon for disambiguation and syntax.** In a CG description of a language the different levels of analysis are closely integrated. Thus, the lexicon should be expressly designed to meet the requirements of disambiguation and syntax. Below we will consider, from this point of view, the treatment of determiners and participles within the SWECG framework. – The background for the considerations is this. When the SWETWOL lexicon was originally compiled (mainly on the basis of *Svenska Akademiens ordlista över svenska språket*, see Karlsson 1992), there was as yet no specific scheme in mind for the syntactic analysis of a number of constructions. When SWETWOL later was put to use in SWECG, with specific analyses in mind, it was evident that some of the category assignments in the original SWETWOL were not wholly appropriate considering the goals of SWECG. Therefore, some SWECG-geared changes were made in the SWETWOL description.

**Desiderata for lexical categorisation**. What is the proper classification of words into part-of-speech categories? How many distinctions should one make, and on the basis of what kind of criteria? Of course, there is no **one** correct answer to such questions. In a sense, for instance, a clas-sification of words into only two categories: inflecting vs. non-inflecting, is a proper classification, but hardly useful for many purposes. Much depends on what the classification is intended to be used for, i.e. on the goals of the overall system. The goals of SWECG are morphological disambi-guation and syntactic (functional) analysis. Therefore, among the desiderata for lexical categori-sation within SWECG, two important ones are the following:

> Desideratum (1). The goal being morphological disambiguation:
> structurally unresolvable lexical ambiguities should be avoided.
> Desideratum (2). The goal being syntactic analysis:
> words with the same syntactic function potential should be assigned to the same category.

The point of (1) is obvious in a framework where the task is to disambiguate words on the basis of grammatical context: First of all, you should see to it that the task is feasible in principle. The motivation for (2) is obvious as well, in any framework where the task is to design rules which associate part-of-speech (or part-of-speech-based) categories with syntactic functions. If words with the same function potential are assigned consistently to the same category, the task of formulating such rules becomes more straightforward, and the rules more general, than would be the case if the words were spread over different categories. – Within SWECG, a practical interpretation of (2) is that the part-of-speech (major syntactic) category of a word is a short-hand for the functions the word may have, consider e.g. the category A as treated in the morphosyntactic mapping component (component (5), section 2 above).

In general, three different kinds of properties of words have been referred to when doing part-of-speech categorisation: morphological (i.e. inflectional), syntactic (i.e. functional or distributional), and semantic. The different criteria do not always point in the same direction (see e.g. Teleman 1997), and, as we know from "traditional" grammar, if you try to give equal credence to all the different criteria in a single system of categorisation, the result is bound to be inconsistent as an embodiment of any one of the three different points of view.

Here, the point of view is that of a syntactic analyser (SWECG). Therefore, syntactic rather than either morphological or semantic criteria are regarded as the ultimate basis for assigning words to lexical categories. Considering the needs of syntax, it is obvious that semantic criteria are not very useful. Morphology is more useful, of course, because classes that can be established on an inflectional basis correlate to a high degree with classes that can be established on the basis of shared syntactic potential. (The two types of classes are not necessarily fully coextensive, however, the latter type of class typically being somewhat more comprehensive.) In any case, for a syntactic analyser, mor-phology is not an end in itself. Ultimately, the categories assumed should be motivated by their usefulness to syntactic analysis. As for SWECG syntax, the categorisation is the more useful the more consistently it adheres to desideratum (2) above.

The less than decisive status of morphological criteria is in fact acknowledged in most part-of-speech systems, traditional or otherwise. For instance, why is it that the Swedish words *forntida* and *varje* have been classified (not, say, as particles) but as an adjective and a pronoun? This is, of course, not because of their (non-existent) inflectional properties, but because of the way they com-bine with other words. Not all words have inflectional properties, whereas all words do have syntactic distribution. If you want to have a unified basis for a grammatically useful classification of **all** the words, you must refer, ultimately, to syntactic criteria. Differences between part-of-speech systems can be seen, from this point of view, as differences in how consistently the necessarily syn-tactic criteria are applied. Another source of difference is the level of classificatory refinement assumed for part-of-speech distinctions.

**Determiners**. We are here concerned with words which, when in pre-head use, typically precede any descriptive adjectives in Swedish noun phrases. The distinction between the two pre-head fields, here called the determiner field and the adjective field, is part of the modern tradition in Swedish grammar. Below, underneath our English terms, are given the terms used by Loman (1956), Thorell (1973), and Teleman (1974). The schema with words in positional groups (P1-P6, cf. below) serves here to illustrate what kind of words we regard as determiners.

| **Determiner field** | **Adjective field** |
|---|---|
| Loman's (1956) "bestämningsfält" | "beskrivningsfält" |
| Thorell's (1973) "pronomenfält" | "adjektivfält" |
| Teleman's (1974) "bestämmarattribut" | "adjektivattribut" |

| (P1 | P2 | P3 | P4 | P5 | P6) |
|---|---|---|---|---|---|
| alla | den | min | mycket | annan | sådan |
| båda | denna | din | få | andra | dylik |
|  | en | deras | många | ena | hurdan |
|  | någon |  | två | första |  |
|  | ingen |  | tre | sista |  |

(Loman divides determiners into six positional groups, P1-P6 above (Loman's semantic terms not given here), the interpretation being that, if a word in one P *is* combinable with a word in another P, then they are combined in the order described, in unmarked use. – For useful notes on the difference between determiners and adjectives see Teleman 1974:124.)

In SWECG, the functional analysis label assumed for the words in the determiner field is @DN>, determiner as premodifier of nominal (cf. Teleman's "bestämmarattribut"). For instance, *en* is @DN> in *en ny fråga*, and *några* is @DN> in *några nya*, and both *dessa* and *dina* are @DN> in *dessa dina etablerade vänner*. Most of these words can also be used as heads. In that use, they get a head function label, e.g. *en* is @OBJ in *Jag tar en*, and *en* is @DN> and *annan* is @OBJ in *Jag tar en annan*.

As for part-of-speech categorisation, the table below summarises first how (some of) these words were treated in the original SWETWOL. On the topmost line are given the part-of-speech tags: ART (article), PRON (pronoun), A (adjective), and NUM (numeral). The '+' in a part-of-speech column indicates that the item in the left-hand column belongs to that part-of-speech; if ambiguous there is a '+' in more than one column.

SWETWOL prior to SWECG:

|  | ART | PRON | A | NUM |
|---|---|---|---|---|
| den | + | + |  |  |
| en | + | + |  | + |
| två |  |  |  | + |
| någon |  | + |  |  |
| alla |  | + |  |  |
| få |  | + |  |  |
| samma |  | + |  |  |
| somliga |  | + |  |  |
| sådana |  | + |  |  |

| | | |
|---|---|---|
| dylika | | + |
| samtliga+ | | + |
| övriga | | + |
| många | + | + |
| andra | + | + |

If we first ignore the first three items in the above table, we can note that the words were catego-rised in three different ways: only as PRON (e.g. *sådana*), only as A (e.g. *dylika*), or as ambiguous between PRON and A (e.g. *många*). Considering that these words are to receive a uniform func-tional analysis in SWECG, it was not optimal that they were categorised in different ways. It is also not obvious how the ambiguity between PRON and A should be resolved, distributionally similar words being categorised unambiguously as either PRON or A.

The treatment of *den* and *en* made for unresolvable ambiguities. Each time when confronted with the word *en* in pre-head use (e.g. *en (ny) fråga*) you had two options, ART and NUM, the one analysis being structurally as feasible as the other one. Similarily, assuming a traditional analysis, for each instance of *den* in construction with adjective and noun (e.g. *den nya frågan*) you were faced with a structurally unresolvable ambiguity, this time between ART and (stressed) PRON. Choosing among options such as these is not something that a grammar-based disambiguator should be forced to try to do. Therefore, the ambiguities should not be introduced in the first place.

The traditional articles may have properties not shared by other determiners (cf. Andersson 1983:294). But the same could be said about many other sub-types of determiners as well, such as possessives, demonstratives, interrogatives, etc. We do not propose that, because of these special properties, the possessives, etc. should be acknowledged as separate part-of-speech level categories. The same applies to articular determiners. (Cf. also Teleman 1997:14.)

The goal of SWECG is functional analysis. Considered from this point of view, the traditional articles do not need to be accorded a separate part-of-speech status in SWECG, the assumed possible functions of *en* and *den* being the same as those of other determiners. This applies also to numerals. For instance, *två* is @DN> in *Jag har två frågor*, and @OBJ in *Jag har två*.

Articles are dispensed with as a separate part-of-speech category also in other frameworks, e.g. *Svenska Akademiens grammatik* (see Teleman 1997), the linguistic annotation system of Ejerhed et al. (1992), the grammar of Andersson (1993), and, it may be added, the Norwegian grammar of Faarlund et al. (1997), and the German grammar of Helbig & Buscha (1991). In the last three, neither are numerals seen as a separate part-of-speech category. In *Svenska Akademiens grammatik* numerals are kept as a separate category, but, according to Teleman (1997:14), only "[e]fter mycket om och men", and mainly because of "trycket från traditionen".

In the SWETWOL used in SWECG, we wanted to avoid the problems or inconsistencies noted above. That is, we wanted to avoid structurally unresolvable ambiguity, and we wanted to provide a consistent lexical category treatment as a basis for consistent function assignment rules. The solution is shown in the table below. The words are categorised as ambiguous between DET (determiner) and PRON (pronoun): when the word is in pre-head use it is to be identified as DET, when used as a head (instead of an N, so to speak) it is to be identified as PRON. As already noted, article (ART) is dispensed with as a major part-of-speech category, and so is the category of numerals (NUM).

SWETWOL in SWECG:

| | DET | PRON |
|---|---|---|
| den | + | + |
| en | + | + |
| två | + | + |
| någon | + | + |

(etc. for all the words in the previous table)

DET and PRON are here the major categories. Any major category can be divided into subcatego-ries. In SWETWOL, any type of subcategory information can be given as angle-bracketed features. As shown below, the DET reading of *en* has the feature <NUM/ART>, and the DET reading of *den* has the feature <DEM/ART>; these features, in fact, serve no practical purpose within SWECG, being included merely to suggest the correspondence to traditional categories.

```
"<en>"
   "en" <NUM/ART> <ID> DET UTR INDEF SG NOM @DN>
   "en" <NUM> PRON UTR INDEF SG NOM
"<den>"
   "den" <DEM/ART> <DF/MD> DET UTR DEF SG NOM @DN>
   "den" <DEM> PRON UTR DEF SG NOM
"<två>"
   "två" <NUM> <ID> DET UTR/NEU DEF/INDEF PL NOM @DN>
   "två" <NUM> PRON UTR/NEU DEF/INDEF PL NOM
"<mina>"
   "min" <POSS-SG1> <MD> DET UTR/NEU DEF PL NOM @DN>
   "min" <POSS-SG1> PRON UTR/NEU DEF PL NOM
```

A determiner that introduces a nuclear noun phrase governs the definiteness forms of the adjective and the noun in the phrase. For instance, *varje* takes both the adjective and the noun in INDEF form, *vilkendera* takes both in DEF form, and *mina* takes the adjective in DEF form and the noun in INDEF form. This syntactic aspect of definiteness is not reflected consistently in the traditional definiteness features of determiners, assigned mainly on the basis of the inherent (semantic) defi-niteness of the words. Therefore, in order to have a reliable basis for syntactic compatibility conditions also in this area, we have introduced a set of features that acknowledge exclusively the syntactic aspect. These features, preceding the part-of-speech tag DET in the readings, we call the governmental definiteness features of determiners; see the Appendix for a list.

**Participles.** The points to be made here are the same in principle as with determiners above. That is, we have made some changes in SWETWOL in order to avoid structurally unresolvable ambigui-ties, and in order to make the lexical description more consistent syntactically.

Participles can be assumed to have the same range of possible functions (or syntactic positions) as "regular" adjectives. The examples (1-6) below illustrate some of the functions: @AN> (1, 2), @SCOMP (3) (cf. Teleman 1974:61), @SUBJ (4), and @ADVL (5, 6). There may be differences of opinion as to the most appropriate functional label in some of the constructions, but the only point made here is that the functions are assumed to be the **same** for the bold-faced words in each pair of examples (1-6).

| Adjective | Participle |
|---|---|
| (1) Ett ***förhatligt***(@AN>) förslag | Ett ***förkastat***(@AN>) förslag |
| (2) De i Finland ***populära***(@AN>) besluten | De i Finland ***fattade***(@AN>) besluten |
| (3) Den har blivit ***färdig***(@SCOMP) | Den har blivit ***accepterad***(@SCOMP) |
| (4) De ***rika***(@SUBJ) stannade kvar | De ***sårade***(@SUBJ) stannade kvar |
| (5) **Trött**(@ADVL) på sitt liv kunde han … | ***Klädd***(@ADVL) i frack kunde han …. |
| (6) ***Överlägsen***(@ADVL) alla andra ville han … | ***Berövad***(@ADVL) sitt pass ville han ... |

An important point here is that, in the description of Swedish, the indeclinable verb-form in present and past perfect tenses, e.g. *dämpat* in *De hade dämpat explosionen*, is treated as a category of its own (V SUPINE).

As for part-of-speech categorisation, the table below summarises first how past participial forms were treated in the original SWETWOL. The part-of-speech tags are V and A, and the feature PCP2 stands for past participle. The abbreviation infl/agr stands for the list of inflection/agreement features, e.g. *berörd* and *bunden* have the features UTR INDEF SG NOM.

SWETWOL prior to SWECG:

| | V PCP2 infl/agr | A infl/agr |
|---|---|---|
| berörd | + | |
| intresserad | + | |
| berörda | + | + |
| bunden | + | + |
| eftersökt | + | + |
| outvecklad | | + |
| vuxen | | + |

The words were categorised in three different ways: only as V (e.g. *berörd*), only as A (e.g. *vuxen*), or as ambiguous between V and A (e.g. *berörda*). Considering that the words are to receive a uni-form functional treatment in SWECG, this categorisation was not appropriate. In particular, the ambiguity between V and A was difficult to make sense of, distributionally similar words being categorised unambiguously as either V or A.

How should past participles be categorised as to part-of-speech membership? In SWECG, on the basis of what has been explained above, the answer is evident: participles are assumed to have the same possible functions as adjectives (and none of the functions of verbs), so participles belong to the category of adjectives. The table below summarises the revised treatment.

SWETWOL in SWECG:

| | | A infl/agr |
|---|---|---|
| Berörd | <PCP2> | + |
| Intresserad | <PCP2> | + |
| Berörda | <PCP2> | + |
| Bunden | <PCP2> | + |
| Eftersökt | <PCP2> | + |
| Outvecklad | | + |
| Vuxen | | + |

The main point of the above table is that all the words belong to A; this is important to SWECG. – Some of the words have the additional, subcategory feature <PCP2>. The descriptive lexical problem concerning this feature is whether a word should be understood as a verb-derived adjective (with the feature <PCP2>), or as a base-form adjective (without <PCP2>). One might want to say that the former is the case when there is a transparent relationship to a corresponding verb, but when is the relationship transparent? We have no answer. In the present description, those words (more or less) have the feature <PCP2> that, in the original SWETWOL, had the feature PCP2 in at least one of their readings. For instance, *berörd* has a reading suggesting it to be verb-derived, *"beröra" <PCP2> A UTR INDEF SG NOM*, and *vuxen* has a reading suggesting it to be a base-form adjective, *"vuxen" A UTR INDEF SG NOM*. – For a word ambiguous between <PCP2> A and V SUPINE see *dämpat* in section 2 above.

Present participles may be argued to be less clearly adjectival in certain respects than past participles (Andersson 1993:284). In SWECG, however, we treat present participles on the analogy of past participles. That is, the function labels we propose to assign to present participles are the same as those we assign to past participles (and other adjectives). Therefore, we categorise present participles as adjectives as well. The subcateory feature is <PCP1>. These word-forms are typically ambiguous between A and N. For instance, the word-form *hårdnande* has the analysis <PCP1> A UTR/NEU

DEF/INDEF SG/PL NOM and also the analysis N NEU INDEF SG NOM. Choosing between the two is a task for the disambiguator.

To sum up, we can refer to Lindroth (1906:10): "Part. [particip] och adj. [adjektiv] öfverens-stämma med afseende på funktion och flexion, men äro olika med afseende på betydelsen". We think too that participles and adjectives are alike morphosyntactically ("med afseende på funktion och flexion"), which is why we assign them to the same part-of-speech category. The semantic interpretation of words is not used as a part-of-speech criterion in SWECG.

## Acknowledgements

At Lingsoft, Pasi Ryhänen and Krister Lindén have provided solutions to innumerable technical problems. Fredrik Westerlund has helped in various ways with the SWETWOL description. At the University of Helsinki, I am indebted to Fred Karlsson and Kari Pitkänen for orientation into and discussion of all aspects of SWECG. For useful comments on an earlier version of this paper I want to thank Krister Lindén, Kari Pitkänen, and Fredrik Westerlund.

*Lingsoft, Inc.*
*Tehtaankatu 27-29 K, FIN-00150 Helsinki*
*jbirn@lingsoft.fi*

## References

Andersson, Erik. 1983. Om gränsen mellan substantiv och nominalfras. In: *Struktur och variation. En festskrift till Bengt Loman*. Åbo, 43–63.

Andersson, Erik. 1993. *Grammatik från grunden*. Uppsala: Hallgren & Fallgren.

Carlberger, Johan & Viggo Kann. 1998. Implementing an efficient part-of-speech tagger. Available at: http://www.nada.kth.se/theory/projects/granska/.

Ejerhed, Eva, Gunnel Källgren, Ola Wennstedt & Magnus Åström. 1992. *The Linguistic Annotation System of the Stockholm–Umeå Corpus Project*. Department of General Linguistics, University of Umeå, Report no. 33.

Faarlund, Jan Terje, Svein Lie & Kjell Ivar Vannebo. 1997. *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.

Helbig, Gerhard & Joachim Buscha. 1991. *Deutsche Grammatik*. Berlin: Langenscheidt.

Karlsson, Fred. 1990. Constraint Grammar as a Framework for Parsing Running Text. In: Karlgren, Hans (ed.), *Papers presented to the 13th International Conference on Computational Linguistics*, Vol. 3. Helsinki, 168–173.

Karlsson, Fred. 1992. SWETWOL: A Comprehensive Morphological Analyser for Swedish. *Nordic Journal of Linguistics* 15, 1–45.

Karlsson, Fred, Atro Voutilainen, Juha Heikkilä & Arto Anttila (eds.). 1995. *Constraint Grammar. A Language-Independent System for Parsing Unrestricted Text*. Berlin and New York: Mouton de Gruyter.

Koskenniemi, Kimmo. 1983. *Two-level Morphology. A General Computational Model for Word-form Recognition and Production*. Department of General Linguistics, University of Helsinki, Publication No. 11.

Lindroth, Hjalmar. 1906. *Om adjektivering af particip.En studie inom nusvensk betydelselära*. Lund.

Loman, Bengt. 1956. Om relationen mellan ordföljd och betydelse hos framförställda attributiva bestämningar till substantiviska huvudord. *Arkiv för nordisk filologi* 71, 218–244.

Mamlmgren, Sven-Göran. 1994. *Svensk lexikografi*. Lund: Studentlitteratur.

Samuelsson, Christer & Atro Voutilainen. 1997. Comparing a Linguistic and a Stochastic Tagger. In: *Proceedings of the 35th Annual Meeting of the Association for Computational Lingustics*. Madrid, 246–253.

Teleman, Ulf. 1974. *Manual för grammatisk beskrivning av talad och skriven svenska*. Lund: Studentlitteratur.

Teleman, Ulf. 1997. Taxonomin i Svenska Akademiens grammatik. Några problem och lösningar. In: Haapamäki, Saara (ed.), *Svenskan i Finland 4*. Skrifter från Svenska institutionen vid Åbo Akademi 3/1997, 9–21.

Thorell, Olof. 1973. *Svensk grammatik*. Stockholm: Esselte Studium.

Voutilainen, Atro. 1995. Morphological disambiguation. In: Karlsson et al. 1995, 165–284.

## Appendix: Grammatical tags used in SWECG

**Part-of-speech**
V = verb, N = noun, PRON = pronoun, DET = determiner, A = adjective, ADV = adverb, PREP = preposition, SC = subordinating conjunction, CC = coordinating conjunction, INFMARK = infinitive marker, INTERJ = interjection, ABBR = abbreviation

**Verbal inflection**
PRES = present tense, PAST = past tense, IMP = imperative, CNJV = conjunctive, INF = infinitive, SUPINE = supine, ACT = active, PASS = passive, DEP = deponential

**Nominal inflection**
UTR = utrum (n-gender), NEU = neuter (t-gender), UTR/NEU = utrum/neuter, UTR-MASC = masculine, DEF = definite, INDEF = indefinite, DEF/INDEF = definite/indefinite, SG = singular, PL = plural, SG/PL = singular/plural, NOM = nominative, ACC = accusative, GEN = genitive, NOM/ACC = nominative/accusative, NOM/GEN = nominative/genitive

**Governmental definiteness features of determiners**
<ID> = indefinite: governs A INDEF and N INDEF, <DF> = definite: governs A DEF and N DEF, <MD> = mixed definite: governs A DEF and N INDEF, <DF/ID> = <DF> or <ID>, <DF/MD> = <DF> or <MD>, <ID/MD> = <ID> or <MD>

**Other additional features** (restricted to tags referred to in this paper)
<NUM> = numeral, <NUM/ART> = numeral/articular, <DEM> = demonstrative, <DEM/ART> = demonstrative/articular, <POSS-SG1> = possessive singular 1st person, <PCP1> = present participle, <PCP2> = past participle, <AUX> = auxiliary, <COLLOCATION> = idiomatic construction or collocation, <NON-SWETWOL> = reading assigned by morphological heuristics, <*c> = upper case represented in the lexicon, <**c> = upper case not represented in the lexicon, CLB = clause boundary

**Syntactic function labels**
@+FMV = finite main verb, @-FMV = non-finite main verb, @+FCV = finite (subject)chain verb, @-FCV = non-finite (subject)chain verb, @SUBJ = subject, @FSUBJ = formal subject, @OBJ = object, @IOBJ = indirect object, @AOBJ = object of adjective, @FOBJ = formal object, @SCOMP = subject complement, @OCOMP = object complement, @ADVL = adverbial, @NPHR = stray nominal head, @DN> = determiner as premodifier of nominal, @AN> = adjective as premodifier of nominal, @GN>

= genitive as premodifier of nominal, @NN> = noun as premodifier of nominal, @AD> = adverbal premodifier, @PA> = preposition as premodifier of adjective, @<P = comple-ment of preposition, @P>> = complement of deferred preposition, @<NN = noun as postmodifier of nominal (apposition), @<N = other postmodifier of nominal, @<ADV = postmodifier of adverb, @<SUBJM = subject marker, @SC = subordinator, @CC = coordinator, @IMCV = infinitive marker with chain verb, @INTERJ = interjection

# BmNorTWOL and NyNorTWOL -
# Two Two-level Models of Norwegian[*]

*Sjur Nørstebø Moshagen*

**Abstract**
This paper describes the two two-level morphological models for Norwegian which have been developed at Lingsoft. These two models cover Norwegian Bokmål and Norwegian Nynorsk, respectively. In the first section, a general overview of the lexicons will be given. Then the description of the lexicons will focus on some of the main strengths of the models, namely the compounding mechanism and the two-level rules. Some of the problems found in free compounding will be described, and a discussion of how these have been resolved will follow. Finally some numbers will be presented concerning the performance of the morphological models.

## 1. Overview[1]

The two-level morphological models (TWOLs) for Norwegian were designed as general morphological descriptions of Norwegian, which in turn can be utilised in different contexts for applications such as spell checkers, hyphenators, indexers, and morphological analysers. The present development is intended more for normative use (such as spelling and grammar checkers and hyphenators) than for descriptive use, but the general design is of course descriptive by its nature: Lingsoft's two-level models are currently being used by many universities throughout the world. The only consequence of the normative focus in these projects is that non-standard spellings of words and abbreviations are not found in the present lexicons, and are thus absent from the models. There is no technical reason why these expressions cannot be added, though, and they might be at a later stage.

The two-level model is a mathematical description of how to turn a set of rules and lexicons into a set of finite-state automata. The model can be seen as a large directed graph, which one traverses when analysing a word (each traverse equalling one analysis), made up of many larger and smaller lexicons pointing to each other. For more details about the two-level model, see Koskenniemi 1983.

The morphological models for Norwegian are organised into three parts: the lexicon file itself; an inflection, compounding and derivation module; and the two-level rules. In the following sections an overview of the lexicon structure, its content, the two-level rules, and the compounding system will be given.

## 1.1 The lexicon structure

The lexicon file is a text file with the information in a tabular set-up as in (1):

| | | | | | |
|---|---|---|---|---|---|
| (1) | bok | bok HOXPLEROML | COMP_NOUN_O | xxnominalxx | 3321 |
| | elv | elv HANXPLERAR | COMP_NOUN_E | xxnominalxx | 7684 |
| | elv | elv HOXPLARER | +COMP_NOUN_E | xxnominalxx | 7683 |
| | elvarme | el#varmE1 HANXPLAR | COMP_NOUN | xxnominalxx | 7685 |
| | elvebåt | elve#båt HANXPLAR | COMP_NOUN | xxnominalxx | 7686 |
| | elveeigar | elve#eigar HANXPLAR | COMP_NOUN_O | xxnominalxx | 7687 |
| | elveigar | elv#eigar HANXPLAR | COMP_NOUN_O | xxnominalxx | 7687 |
| | fadder | fadD2E2r HANXPLAR | COMP_NOUN_O | xxnominalxx | 8425 |
| | hogga | hogD2 STERKEX | COMP_VERB_E | xxverbalxx | stv153_hogga |
| | skål | skål HOXPLER | COMP_NOUN | xxnominalxx | 43717 |

The table has six fields:

- dictionary entry form
- lexicon entry form
- inflection code
- compounding and derivation code
- lexicon class

- ID (only in Nynorsk)

The leftmost column is used both as an aid for the linguist (some of the lexicon forms of words can be quite hard to read), and as the basis for restricting compounding according to word length (see section 4). The next field, the lexicon entry field, is the actual entry in the morphological model. The symbols used to write the entries are those defined as lexical symbols by the two-level rules. Thus some of the entries can appear cryptic (e.g., *fadD2E2r* for the lexeme 'fadder' (*godfather, -mother*), *kJ2o[U2s* for the lexeme 'kjosa' (*choose, select*); see section 2 for a short explanation of some of these symbols), but most of the symbols are normal alphabetic symbols. Following the lexicon entry field, the inflection field names the continuation lexicon that defines the inflectional behaviour of a lexeme. The next field is the compounding field, which names the continuation lexicon defining the compounding behaviour. The named continuation lexicon describes what combining segment(s) a word takes, if any, and it also describes the set of suffixes the word can take. To a certain degree the compounding lexicon also defines which kind of words it can compound with. The last field used in the present model is the lexicon class field. This field determines the main lexicon to which a word belongs, roughly equalling the base parts-of-speech. It is a technical definition, though, and to a certain degree this field is also used for regulating compounding. The ID field is not used for anything at present, except for identifying the lines of entries belonging to the same lexeme (one may think of this as a lexeme identifier). Also, this field is only used in Nynorsk.

At compilation time, this table is transformed into a set of lexicons of the form required by the TWOL compiler. The use of a tabular set-up makes it quite easy to edit and maintain the main body of the lexicon (and the model), and gives a very flexible way of adjusting compounding and inflection. This same format is also used by the Danish TWOL.

### 1.2 Lexicon content

The content of the lexicons is as follows[2]:

| Word type | Bokmål | Nynorsk |
|---|---|---|
| Verbs | 7 449 | 7 262 |
| Nouns | 43 165 | 40 733 |
| Adjectives | 11 997 | 10 853 |
| Uninflectable words | 2 004 | 1 978 |
| Proper names | 12 618 | 12 853 |
| Geographical names | 7 690 | 7 879 |
| Personal names | 3 864 | 3 865 |
| Corporate names | 403 | 418 |
| Abbreviations | 835 | 790 |
| Totals | 74 499 | 73 694 |

The numbers presented in this table might not be 100 % exact, although they are nearly so; many words have several entries for different purposes. As far as possible, such multi-entry lexemes have been counted as one, but some may have slipped through[3]. – The proper names lexicon is shared between the Nynorsk and Bokmål TWOLs, since most names are the same in both versions. There are some differences, though, which are reflected in the numbers. The Bokmål lexicon is based on Langfald og Paulssen 1996, whereas the Nynorsk lexicon is based on Nynorskordboka 1994, with updates as of March 1998.

## 2. Two-level rules

The two-level model includes a module for writing rules to describe correspondences between lexical and surface forms. They were originally thought of as phonological rules; in this case, though, they are applied to written language. The Nynorsk model has 37 such rules, and Bokmål 25, some of which will be described below.

The two-level rules are used to express regular changes in words, typically changes that occur when the word is inflected. By expressing such changes as 'phonological' rules, the number of inflectional classes can be greatly reduced. In (2) are given a simplified description of such a rule, the rule stating that a set of consonants can double in certain contexts. The rule in the two-level formalism is shown in (3).

(2) Consonant doubling:

```
b–>bb
d–>dd
f–>ff      / __ { V | # | derivational ending }
g–>gg
...
```

(3) "consonant doubling allowed"

```
0:X <=> X: _ D2: ( TempUmlDia ) [ %~: | ( %>: )  Vowel | ( %\: ) #: ] ;
            where X in DobCons ;
```

Another rule describing vowel deletion in certain contexts is shown in (4):

(4)

```
E2:0 <=> _ Likvida ( %°: ) [ Plural1 | Plural2 | Plural3 |  %~: Vowel ] ;

E2:e <=> _ Likvida ( [ DefMasc | DefFem ] ) ( Genitiv ) ( %\: ( :s ) ) #: ;
```

These two rules work together to give the vowel and consonant alternations in a word like 'fadder' (godfather, -mother), as seen in (5); the lexicon entry of this word was listed in (1) above:

(5) fadder

fadderen

fadrar

fadrane

The morphological and phonological descriptions used as a basis for the TWOLs are taken from Beito 1986, and Faarlund et al. 1996.

## 3. Compounding

As mentioned earlier, compounding is an important part of the description of Norwegian. People use compounding all the time, creating new words 'on the fly' when talking or writing. A morphological model of Norwegian (and many other languages as well) without a proper treatment of compounding will be severely hampered (that is not to say that the following way of doing it is the only one, or the best one). In (6-9), some compounds found in our corpus are listed, as examples of the creative use of compounding.

(6) normal compounds:

arbeidsløysesamfunnet (*the unemployment society*)

avisteikneserieland (*newspaper cartoon land*)

jesusrevolusjonære (*Jesus revolutionaries*)

overkompensasjonsregel (*over-compensation rule*)

overpensjoneringsproblem (*over-retirement problem*)

samfunnsminnefunksjonen (*the society memory function*)

(7) compounds with names:

Aust-Timor-saka (*the East Timor case*)

Nome-gullgravar (*gold miner from/in Nome*)

(8) compounds with abbreviations:

EØS-parlamentarikarkomiteen (*the committee of EEC parliamentarians*)

SV-utbrytarar (*SV breakaway*)

(9) compounds with numbers:

1890-årsmodernistar (*1890's modernists*)

Compounding in the two-level model (TWOL) is generally done the same way as inflection, through continuation lexicons. But where inflectional lexicons are final (they all end in a termination symbol), compounding lexicons point back to the root of the tree (or beginning of the graph), thus making it possible to traverse the graph an indefinite number of times, in effect making possible indefinitely long compounds. It also makes possible combining all words with one another, thus covering all the possible and thinkable (indeed, even unthinkable) compounds covered by the lexicons. Some of the compounds that can be morphologically described may be 'semantically challenging', but it is not the job of morphology to consider semantics.

From a descriptive point of view, the goal of the compounding system is to describe as accurately as possible the compounding behaviour of each word (or lexeme). That is done by specifying for each word which of a set of compounding lexicons it can take. The compounding lexicon specifies whether or not to use a 'filler' segment[4], and if so, what segment to use, where to go from there (that is, back to the root), and also what derivational endings or suffixes can be combined with words using a specific compounding lexicon.

In most cases this gives a good and accurate description of a word's behaviour. But the present model is too coarse to give the necessary control over the details in compounding. There are two problem areas in particular. The first is the specification of the filler segment, which may vary depending on the position of the word in a compound, and also depending on its relation to the surrounding compounding parts. The classic example is given in (10), where the word 'skinn' can take a filling segment or not, depending on different positions and relations of the word in the compound; some other examples are given in (11—12)[5]:

| | | | |
|---|---|---|---|
| (10) | **skinn**hanske | -0- | (leather glove) |
| | dame**skinns**hanske | -s- | (leather glove made of a woman's skin) |
| | dame**skinn**hanske | -0- | (women's leather glove) |

|  |  |  |  |
|---|---|---|---|
| (11) | **jul**ebrød | -e- | (Christmas' bread) |
|  | rom**jul**sbrød | -s- | (bread for the days between Christmas and New Year's eve) |
| (12) | **råd**smøte | -s- | (meeting) |
|  | **råd**gjevar | -0- | (adviser) |
|  | u**råd**sgjevar | -s- | (adviser giving bad advice) |
|  | heime**råd**velde | -0- | (home rule) |
|  | heime**råd**svelde | -s- | (rule by a family council) |

The present model does not allow us to specify filling segment variation to the extent shown in (10–12). To try to minimise the errors, three different strategies have been used. None of the strategies does the whole job, but together they manage fairly well.

As a first attempt to catch the filling segment 's' very often used in the second border, a special two-level rule has been designed that allows a special filler 's' to show up only if preceded by at least one word boundary. This will over-generate, but it won't miss any compounds such as have been described. The second way of handling this variation is to specify several different filling elements for one compounding lexicon. This will probably also over-generate, but to a lesser extent than the first one, provided the information in the lexicon is accurate. The third way of increasing the precision is to lexicalise as many compounds as possible, thus escaping the whole problem. If one can lexicalise all compounds of an (almost) non-productive compounding pattern for a word, one can remove the corresponding filling segment from the compounding specification, and thus reduce the over-generation and increase precision.

The other main problem with the present model of compounding is that compounding and derivation are seen as one and the same thing from a technical point of view in the TWOL: when you specify compounding, you also specify derivation. For big groups of words this may be true, but it is certainly not true for all words. One simple example could be the derivational suffixes used for Latin- and Greek-based loan words, and those used for Norwegian words. Although many of the loan words and the 'native' words both compound with the same filler segment, and continue to the same classes of words, they share very few derivational suffixes.

## 4. Restricting compounding

Free compounding is desirable and necessary in the model, since it is a feature of the languages to be described. But it has its price. The single most important aspect in the context of this project is precision. Since the first use of the model will be in prescriptive/normative applications, it is important that it will not over-generate, that is, that it will not analyse misspellings as some kind of compound. At the same time, it has been one of the goals of this development project to keep the compounding system as open and flexible as possible.

The main mechanism used for controlling compounding has been a length-based division of words into two groups: those which can compound, and those which can't. The limit has been set mechanically to four characters; that is, all words shorter than four characters are excepted from compounding. This is done through a parameter in the compilation script, so that this limit can be reset to other values (including no limit at all) for purposes where compounding restrictions are unnecessary or unwanted. On top of that a system for turning on or off compounding of individual words has been implemented, so that the setting can override the basic length-based division. Thus, words shorter than four characters can have the compounding function turned on, and longer words can have it turned off. The basic criteria for turning compounding on or off for individual words has been an evaluation of how much is gained and lost in recognition and analysis power in each case, and how that influences the acceptance of wrongly spelled words. The basic assumption behind this is that shorter segments are more likely to be found in erroneous analyses than longer ones, and by restricting the short segments, those problems can be avoided.

The power lost because of this restriction is shown in the analysis (or rejection) of short compounds, that is, compounds composed of at least one such short segment; these words remain unanalysed or have lost the analyses involving the short segment. The gap created by the restrictions has been filled as far as possible by lexicalising the missing compounds, and by letting the lexicalised ones take part in compounding in normal fashion.

These limitations are arbitrary technical restrictions which at any time can be removed. Especially for analysis purposes, it is more important to give all words at least one correct analysis than toremove many wrong analyses. On the other hand, by having fewer analyses to consider, the burden on a disambiguator in, for example, a syntactic analysis context, could be reduced.

Some examples of words which were accepted in an earlier version are given below in (13) and (14), together with the wrong analysis and a short explanation of what has been done (both examples are Nynorsk):

| | | | |
|---|---|---|---|
| (13) | fortellingar | correct: | forteljingar (= *stories*) |
| | fort\#ell~ing | action: | the verb 'elle' has been removed from compounding, since it (and its derivations) has the potential of being mixed-up with many misspellings, and it is not found in any compound. |
| (14) | avholde | correct: | avhalde (or: halde av) (= *be fond of*) |
| | av\|holde | action: | the verb 'holde', which in Nynorsk means something like '*get flesh on the bones*', is removed from compounding (and therefore also derivation), to reduce the likelihood of being mixed with the Bokmål word 'holde' in, e.g., compounds and derivations. |

These examples show some of the considerations taken when developing the described TWOL.

## 5. Performance

In this section I will present some performance numbers for the two TWOLs. First there is a short description of some indicative speed numbers, then a more thorough discussion of the quality of the TWOLs. Only numbers for Nynorsk are given in this discussion, but the results should be indicative for the Bokmål TWOL as well.

The test material used in both cases is a sorted word list consisting of unique word forms. For Nynorsk the list contains 99,601 unique word forms, roughly equalling 500,000 words of running text. The Bokmål list contains 226 661 unique word forms, roughly 900,000 words of running text. The lists are the same as the ones used for testing during development; thus the results will be biased. The development testing has only been testing of the differences between older and newer versions of the TWOLs, and has never been a source for lexicalisation. The bias should therefore not be too disturbing.

The present TWOLs do not contain any pre-processor or mechanisms for handling capitalised words. That is the main reason for the low recognition rate, slightly above 50 %[6].

### 5.1 Speed

Speed numbers will of course depend on the computer used, the operating system load, and other factors. The numbers presented here should therefore be seen as indicative only, and not as absolute figures. The test was quite simple, running a word list of unique word forms through the analyser, and measuring the time taken. The test was run on an Intel Pentium II 300 MHz system, running Linux.

Speed numbers (# of words analysed pr. second):

Bokmål: 226,661 words analysed (124,389 received an analysis)

in 314.54 seconds = 720.61 words / second

Nynorsk: 99,601 words analysed (51,752 received an analysis)

in 355.52 seconds = 280.16 words / second

The speed tests were run using a version of the TWOLs that was unoptimised. An optimised version of the TWOLs would be faster by 5-10 % (and take less disk space).

The most striking result is the big difference between NyNorTWOL and BmNorTWOL, the Bokmål TWOL being more than twice as fast. There are of course differences between them, both in the number of two-level rules, and in the construction of the lexicon. But those differences should give rise to such big differences in speed. There has been no time to investigate this further, but similar results have been reported from Taggerprosjektet at the University of Oslo[7].

### 5.2 Quality of the Analyses

The quality of a morphological analyser can be measured in several different ways. In the following discussion I will focus on two different aspects: the internal analysis of a word (boundary identification; base form identification) and the morphosyntactic analysis of a word. Because of the lack of a preprocessor or similar device to handle capitalised words, it is presently meaningless to look at the words *not* recognised. Thus I will focus on the words recognised. It is further important to emphasise that the following test and evaluation is done with the version of the TWOL described in this paper, that is, a TWOL with the compounding restrictions described in section 4. Only the NyNorTWOL is tested and evaluated, but the results should be similar for BmNorTWOL.

The task of the test and evaluation is to identify the number of errors made by the TWOL, the type of error, and its cause. As mentioned earlier, a word list consisting of 99,601 unique word forms was run through the analyser. 51,752 of the words were recognised (see above about capitalised words; also unknown names, foreign words and unrecognised compounds due to the compounding restrictions belong here). 22,844 words were analysed due to the compounding mechanism[8], the rest analysed as non-compounds, or lexicalised compounds. 'Analysed as' is to be understood as: among all the analyses given to a word, at least one of them is correct. A wrongly analysed word is a word which does not have *any* correct analysis.

Since the likelihood of making mistakes when compounding two known elements is quite low, I decided to disregard such words from further analysis, assuming they are all correctly analysed. And all non-compound words are by definition correctly analysed. The remaining words are the interesting ones: compound words consisting of three or more compound elements (in the technical sense of the TWOL). These are the ones most likely to contain parser errors, and eventually be analysed without any correct analysis at all.

1069 such compounds were found, and inspected manually for errors. 153 words were identified as having *no* analysis that was fully correct. These words were further inspected, to classify the errors found.

The first classification was based on the internal analysis of the words, that is, word boundary identification and base form identification. The words were divided into two groups: mostly correct (85), and totally wrong (68). The 'mostly correct' words had only one error: either one wrong baseform (compound lexeme), or one wrong boundary. The other group had more errors. For the 'mostly correct' words, the error might or might not influence the morphosyntactic analysis of the word.

The second classification of the errors was based on the morphosyntactic analyses given, that is, whether they contained at least one morphosyntactic analysis, despite some error in the analysis of a word. The table in (15) gives an overview of the numbers found:

(15)

|  |  |  |
|---|---|---|
|  | 89 | Correct analysis |
| 123 | 14 | Correct parts-of-speech, one or more other errors |
|  | 20 | correct parts-of-speech, one or more other errors, inference from Bokmål |
| 26 | 17 | Totally wrong |
|  | 9 | Totally wrong, inference from Bokmål |
|  | 4 | Unclassifiable |

The four unclassifiable words were one Swedish word, two English words, and one word which was impossible to identify due to missing context ('minnetapetsyte', analysed as a noun or an adjective).

In most cases, the part-of-speech information is the most important one. I therefore further analysed the group of 26 words that did not get anything correct. In (16) an overview of the results is given. The words are here subdivided into two groups, based on the first classification (word internal analysis):

(16) 'one error internal': 8
- 6 words were accidental compounds[9]
- 2 words were Bokmål words

'more errors int.': 18
- 7 words were accidental compounds
- 5 words were Bokmål words

- 3 words were accidental compounds of Bokmål words
- 2 were correct words that did not get a correct analysis
- 1 was a typo that got an accidental analysis

The two correct words that did not get any correct analyses were missed because of compounding restrictions ('ulveskinnsfell', missed because 'fell' has been restricted), and because of a missing entry in the lexicon ('næringslivsturne', missed because 'turné' with an unaccented 'e' in the indefinite singular form was missing from the lexicon).

The results displayed above show that the number of wrong analyses are quite low, despite the restrictions in the compounding mechanisms in terms of part-of-speech only 2 in 50,000 words. Without compounding restrictions, there would have been one error, and that one can be removed by adding the missing word to the lexicon.The results also show that the precision in identifying word boundaries and base forms in compounds is so high that reliable statistical studies of compounding and word formation are now possible. Also here reliability should increase to almost 100 % by removing the compounding restrictions.

*Lingsoft, Inc.*
*Helsinki*

**References**

Beito, Olav T. 1986. Nynorsk grammatikk – Lyd og ordlære. Oslo: Det Norske Samlaget.

Faarlund, Jan Terje, Svein Lie og Kjell Ivar Vannebo 1997. Norsk referansegrammatikk. Oslo : Universitetsforlaget.

Karlsson, Fred 1992. SWETWOL: A Comprehensive Morphological Analyser for Swedish. Nordic Journal of Linguistics, 15, 1-45.

Koskenniemi, Kimmo 1983. TWO-LEVEL MORPHOLOGY: A General Computational Model for Word-Form Recognition and Production. Department of General Linguistics, University of Helsinki, Publication No. 11.

Langfald, Aagot og Kjell M. Paulssen 1996. Norsk ordbok – bokmål. Oslo : J. W. Cappelens Forlag.

Nynorskordboka 1994. Oslo: Det Norske Samlaget.

---

[*] Thanks to Antti Arppe, Thomas Bilgram, Tom Lewis and Fredrik Westerlund for comments and help.

[1] This article describes the two-level morphological models from a rather technical point of view, focusing on numbers and architecture. For a more thorough description of the output from such models, see Karlsson 1992.

[2] The numbers may not add up to the totals given. It has been hard to calculate precise numbers especially in Bokmål,, since there is no ID to identify a set of entries belonging to the same lexeme. The numbers do, however, give an impression of the size of the lexicons, and their different parts.

[3] Because of the use of an ID field in the Nynorsk lexicon, it has been easier to count multi-entry lexemes as one for Nynorsk. The numbers for Nynorsk should therefore be very close to 100% accurate. An exception is the names lexicon – the shared name lexicon is also without an ID field. But very few names have multiple entries.

[4] Such segments could as well be called connectors; historically, they are all genitives, but syncronically they can best be described as a segment whose presence is obligatory between words – in some contexts for some, but not all, words – when making compounds.

[5] Not all example words given are documented words; a native speaker will, however, know when to use what filler segments, also in new and never-before-seen compounds, and the examples given are merely meant to illustrate variation in the use of such fillers.

[6] A mechanism for handling such cases has now been added, but was not ready for testing in time for this article. The recognition rate has risen, though, to almost 2/3 (63.4%, or 63,141 of 99,601). This mechanism has so far only been implemented for Nynorsk.

[7] Oral communication by Anders Nøklestad.

[8] That is, 44.1%.

[9] Words that should not have been compounded - that is, they should have been written as separate words.

# FUNCTIONAL/COGNITIVE LINGUISTICS

# The Subject is Nominative! On Obsolete Axioms and their Deep-Rootedness

*Jóhanna Bar∂dal*

**Abstract**

This paper deals with some of the methodological problems that arise when investigating the potential subjecthood of Oblique Experiencers in Old Scandinavian. It turns out that most scholars adressing this issue have reached conclusions which are in accordance with their premises, a fact which undoubtedly lowers the reliability of their results. Ideology legitimates every methodology, thereby it should be its scientific consequences which ultimately determine our choice of methods.*

## 1. Introduction

At least for the last two decades or so, scholars have debated on the impersonal construction of the older Germanic dialects, and amongst other things, the syntactic status of the Oblique Experiencer as illustrated in the following examples from Modern Icelandic:

(1) a. **Mér** er kalt!
I (dat.) am/is cold
'I am freezing'
b. **Honum** finnst íslenskur hákarl gó∂ur!
He (dat.) finds Icelandic shark (nom.) good
'He likes Icelandic shark'

The impersonal construction has survived only in three modern Germanic languages; Icelandic, Faroese and German. In Modern Icelandic, the Oblique Experiencer passes all subjecthood tests (see Zaenen et al. 1985 and Sigur∂sson

1989, 1992), and the same holds for the Oblique Experiencer in Modern Faroese (see Barnes 1986). In these languages, the Oblique Experiencer has to be analysed as the syntactic subject of the clause. In Modern High German the Oblique Experiencer is not considered a syntactic subject but a syntactic object (see Seefranz-Montag 1983, 1984, Smith 1994, Zaenen et al. 1985), since it does not pass the subjecthood tests of Modern High German. The Oblique Experiencers pass some of the subjecthood tests, though, which has been taken as an indication of them having had the syntactic status of subjects in earlier periods of German (see Seefranz-Montag 1983:165-167). Seefranz-Montag observes that Oblique Experiencers certainly do not behave particularly object-like; she also points out that the impersonal construction is not a part of the everyday, spoken language, it exists almost only in the formal written language and is considered as a frozen relic of older periods of German (1983:200). As I pointed out in Barðdal(1997:47), it is not at all strange that lexicalized relics do not pass tests made to measure productive, syntactic behaviour, as subjecthood indisputably is. Consequently, the fact that Oblique Experiencers do not pass all subjecthood tests of Modern High German does not say anything about their syntactic status in other stages of German.

Even though the syntactic status of the Oblique Experiencers in the modern Germanic languages is quite clear, at least for Icelandic and Faroese, the same is not true for the older periods of the Germanic languages. Scholars basically divide into two camps; those who consider the Oblique Experiencer to be syntactic objects and those who consider them syntactic subjects. The former group includes Butler (1977), and Smith (1994, 1996) for Old English and/or Old High German, Faarlund (1990), Kristoffersen (1991, 1994), Mørck (1992) and Smith (1996) for Old Norse, and Sundman (1986) and Falk (1995, 1997) for Old Swedish. In the latter group we find Harris (1973 (here cited by Butler 1977)), Elmer (1981), Seefranz-Montag (1983, 1984) and Allen (1986, 1995) for Old English and/or Old High German, Rögnvaldsson (1991, 1996) for Old Icelandic and, Barðdal(1997) for Old Scandinavian in general. I have excluded those scholars who have treated the impersonal construction without taking a stand on the syntactic status of the Oblique Experiencers, and those who have taken the traditional stand that they are syntactic objects, such as Fischer and van der Leek (1983, 1987), Denison (1990) and Traugott (1992), to mention only a few.

In a previous study on the subjecthood properties and syntactic status of Oblique Experiencers in Old Scandinavian, I noted that the methodological differences of some of the more recent work on the topic, are huge (Barðdal1997:26-28). These differences are definitely worth a closer inspection, a topic to which the remainder of the paper will be devoted. For reasons of space I will only review the most recent studies and concentrate on those published during the last decade.

I will start by giving a brief summary of my own methodological standpoint, which will then be measured against Faarlund's (1990), Kristoffersen's (1991, 1994), Mørck's (1992) and Falk's (1995, 1997), with a note on some of their particularities and peculiarities (section 2). Then I will summarize and evaluate the most eminent methodological features of the approaches cited above (section 2.6) and close the discussion with some remarks on one particular methodological problem that arises in all work on historical linguistics, namely the problem how our premises affect our results (section 3). The last section is a summary.

## 2. Methodological Differences

### 2.1 Bar∂dal (1997)

In a paper on Oblique Subjects in Old Scandinavian (1997), I have argued that the most reasonable way of viewing constructions in older periods of our languages is in relation to the same constructions in our languages today, as long as they exist. To illustrate my point, let's look at some constructions from Modern Icelandic and compare them with the equivalent constructions in Old Icelandic or Old Scandinavian in general:

(2) a. Ég hjálpa flér
I (nom.) help you (dat.)
'I help you'
b. Mig dreymir Gu∂mundur
I (acc.) dream Gu∂mundurur (acc.)
'I dream about Gu∂mundurur'

*hjálpa* in (2a) selects a nominative subject and a dative object in Modern Icelandic. I argue that if we find an example such as (2a) in Old Icelandic, we can assume that things are the same as in Modern Icelandic. For the sake of consistency, I argue, we should treat the impersonal construction in the same way as the ditransitive. *dreyma* in (2b) selects an accusative experiencer as the subject and an accusative theme as an object. *dreyma* was constructed in the same way in Old Icelandic and as long as there are no reasons for treating the impersonal construction differently in Old Icelandic than in Modern Icelandic, we should also treat them the same. My null hypothesis, following Rögnvaldsson (1996:54-55), is that we only assume a different analysis for constructions that look alike if we have an independent reason for doing so. Further, I claim that this argumentation can be extended to cover all the Old Scandinavian languages and even Old Germanic in general.

Another hypothesis of mine is that we can expect languages which have a case system and a structure similar to Modern Icelandic to also have Oblique Subjects, as Modern Icelandic does.[1] In fact, it turns out that Oblique Experiencers of Old Scandinavian and Old English behave syntactically like Oblique Subjects in Modern Icelandic. Of course not all subjecthood tests for Modern Icelandic are valid subjecthood tests for the earlier periods of the Germanic languages, but some of them are (see Barðdal1997:46):

**Table 1.**

| | MIce | OIce | OSw | ODa | OEng | |
|---|---|---|---|---|---|---|
| Clause-bounded Reflexivization | | xo | x | x | x | x |
| Conjunction Reduction | xo | x | x | x | xo | |
| Word order in decl. main clauses | | xo | x | x | x | x |
| Subject-verb inversion | xo | x | x | x | x | |
| Position in sub. clauses | xo | xo | x | | | |
| Subject-to-object raising | xo | xo | xo | xo | xo | |
| Subject-to-subject raising | xo | xo | xo | xo | xo | |
| PRO-infinitives | xo | xo | xo | | xo | |
| Long distance reflexivization | xo | xo | | | | |

Notes: xo = conclusive subjecthood test, x = non-conclusive subjecthood test but Oblique Experiencers behave like Oblique subjects in Modern Icelandic.

As is obvious from *Table 1*, Oblique Experiencers behave in Old Scandinavian approximately as they do in Modern Icelandic. If we concentrate on Old Icelandic, which shows the highest correlation with Modern Icelandic (which of course could be due to Old Icelandic having been the subject of more research than the other languages), *Table 1* shows that Oblique Experiencers behave exactly like Oblique Subjects in Modern Icelandic, that they occur in constructions containing Clause-bounded reflexivization, Conjunction reduction, with the subject in first position in declarative main clauses, inverted word order, with the subject in first position in subordinate clauses, subject-to-object raising, subject-to-subject raising, PRO-infinitives and even long distance reflexivization. Of these, the last five are conclusive subjecthood tests for Old Icelandic, three are conclusive tests for Old Swedish, two for Old Danish and four for Old English. At the same time it is clear that not all subjecthood tests for the modern languages are real subjecthood tests for the Middle Ages. Clause-bounded reflexivization and Conjunction reduction are not subjecthood tests in Old Scandinavian, since objects could undergo these processes as well. For examples and a more extensive discussion I refer the reader to Bar∂dal(1997). For further research on the development of the impersonal construction in the Mainland Scandinavian languages, based on the idea that Oblique Experiencers are syntactic subjects, I refer the reader to Bar∂dal(1998).

## 2.2 Faarlund (1990)

In his 1990 study on Old Norse, Faarlund elaborates on the following subjecthood tests: position, verb agreement, reflexivization, conjunction reduction, raising infinitives, subject control in infinitive clauses and subject deletion in imperatives.[2] He finds that none of these can be used as a criterion, apart from PRO-infinitives, since either the rules are different from Modern Norwegian, or they do not apply only to nominatives but also to Oblique Experiencers. Since only nominatives are found in PRO-infinitives, Faarlund concludes that only nominatives have subjecthood properties and at the same time, that Old Norse does not have a syntactic subject. This is in accordance with his previous conclusions that Old Norse is non-configurational, while Modern Norwegian is configurational (1990:110).

I will take issue with Faarlund (1990) on several points. First, it is not at all clear that Old Norse is non-configurational. Rögnvaldsson (1995) rejects that claim and argues that even though word order in Old Icelandic is less rigid than in Modern Icelandic, and even though it is possible to find instances of discontinuous phrases, tests for constituency render the same results for Old and Modern Icelandic (1995:21).

Second, Faarlund bases his conclusions that Oblique Experiencers are not syntactic subjects on the conclusion that the subject does not exist as a syntactic category, a conclusion, which again, is based on the fact that nominatives are not distinct from Oblique Experiencers. Thus, Faarlund's conclusions are highly circular and as such unreliable.

Thirdly, it seems as if Faarlund has defined the subject a priori as being the nominative NP and thus by definition excludes Oblique Experiencers. Every test which renders positive results for both nominatives and Oblique Experiencers is systematically rejected as a test for subjects, leaving behind only tests that nominatives pass. In addition, Faarlund uses the terms *subject* and *nominative* as synonyms (1990:125). Considering this, it is not at all surprising that Faarlund reaches the conclusion that Oblique Experiencers are not syntactic subjects.

## 2.3 Kristoffersen (1991, 1994)

In his discussion of syntactic functions, Kristoffersen (1991:70-71) assumes a general definition of the subject, based on Dyvik (1986), which says that the subject is assigned a semantic role by the verb, and that the subject is the only syntactic function which can have the semantic role of agent. Kristoffersen concludes that the nominative in Old Norse satisfies both of these requirements and is therefore a likely subject candidate (p.72). In order to test the hypothesis that the nominative is the subject, he illustrates that the nominative NP is equipped with some of the properties Keenan (1976) assumes for subjects. Kristoffersen (1991) uses the following subjecthood criteria for Old Norse: agreement, position, reflexivization and PRO-infinitives (1991:72-85). And it turns out that the nominative passes all the conclusive tests.

An immediate problem in the logical deduction arises: that nominatives pass the subjecthood tests does not exclude everything else from subjecthood. However, that is nevertheless how Kristoffersen argues. Another problem is that Kristoffersen uses agreement as a subjecthood criterion, thereby excluding everything but nominative. Thus, Kristoffersen has excluded Oblique Subjects, by definition.

Kristoffersen defines objects in the following way (1991:94):

(3) Eit objekt er eit nominalt ledd som ikkje er subjekt, og som har akkusativ, dativ eller genitiv kasus.

'An object is a nominal constituent which is not the subject and which is in the accusative, dative or genitive case.'

This is not an independent definition of object. Consequently, if the definition of subject is flawed in any way, the object is not defined satisfactorily either. Also, this definition of object excludes nominatives (which are found as objects in Modern Icelandic, and in Old Swedish according to Falk 1997:40), not with reference to independent evidence, but rather as a matter of definition.

Kristoffersen concludes (1991:89) that Oblique Experiencers are not subjects since they share too few properties with Oblique subjects in Modern Icelandic. Now, about this, I can only say that since Kristoffersen defines the subject as controlling verb agreement, Oblique subjects are ruled out a priori, and by circular reasoning, the Oblique Experiencer is defined as an object.[3]

In a 1994 paper on the passive in Old Norse and Modern Icelandic, Kristoffersen again raises the question of whether Old Norse had Oblique subjects. Kristoffersen concludes that the evidence for the subjecthood of Oblique Experiencers is too meagre, and therefore he assumes that these are objects. Accordingly, passives of verbs that do not select accusative objects are subjectless. In my opinion, the foundation for analysing these passives as subjectless is very weak. Kristoffersen does not present any evidence that Oblique Experiencers *are not* subjects, he merely illustrates that it is possible to argue against the claim that they are.

## 2.4 Mørck (1992)

After considerable speculation on the nature of the subject, Mørck (1992) reaches the conclusion that a subject should be a homogeneous category, so he argues that nominatives are subjects. Consider the following quote (Mørck:1992:88):

(4) Men vi kan også si at de egenskapene som fins hos mer enn **én leddtype**, ikke kan være subjektsegenskaper. Det vil føre til at heller ikke potensielle oblike subjekt får status som skikkelige subjekt,og vi ender opp med bare nominativledd i denne funksjonen. [my bolding]

'But we can also say that the properties found with more than **one** constituent type can not be subject properties. That will lead us to the conclusion that potential oblique subjects can not be considered proper subjects, and we end up with only nominative constituents having that function.'

I would like to point out that if Oblique Experiencers are syntactic subjects, then they constitute the same clause constituent as nominative subjects, and are not another type of constituent, as Mørck apparently assumes. By assuming a priori that nominatives and Oblique Experiencers are two distinct syntactic constituents one is justified to treat them differently. Mørck decides that nominatives are subjects and thereby Oblique Subjects are excluded, by definition.

The reason Mørck assumes that Oblique Experiencers are another constituent type than nominatives is, apart from nominative case and verb agreement, that they are not found in imperatives, passives and Control Constructions. They are not found in imperatives or passives[4] in Modern Icelandic, and they are very rare in Control Constructions (Rögnvaldsson 1991:372 and 1996:50), yet they are syntactic subjects in Modern Icelandic. This is therefore not a valid counter argument to the claim that Old Norse has Oblique subjects.

Mørck assumes that nominatives are a homogeneous class, or perhaps that verbs selecting nominatives are a homogeneous class. We only have to look at the modern languages to see that subjects behave differently depending on which verbs they are subjects to. In example (5) we see that not all verbs with nominative subjects can occur in passive, in spite of the subjects being nominative:

(5) a.i Hann keypti bókina.
he (nom.) bought book-the (acc.)
'He bought the book'
a.ii Bókin var keypt.
book-the (nom.) was bought
'The book was bought'

b.i Bókin kostaði mikla peninga
book-the (nom.) cost much money (acc.)
'The book cost a lot of money'
b.ii *Miklir peningar/*Mikla peninga var kostað
much money (nom.)/ (acc.) was cost
'A lot of money was cost'

Therefore we can perhaps not expect all nominative subjects to have all subject properties, anyway. Thus, Mørck has presumably not gained anything by defining the subject as nominative, after all.

Another problem with a morphological definition of the subject in Old Norse is that it does not make the category a homogeneous category anyway.

Some nominatives behave like objects and Oblique Experiencers behave like subjects. Consider the following example (from Rögnvaldsson 1996:60) in which the Oblique Experiencer of the verb *flykja* occurs as a PRO in the position in front of the infinitive, while the nominative behaves like an object since it is in the position following the infinitive and does not appear as PRO.

(6) fiór∂ur kva∂st flykja tvennir kostir til
fiór∂ur (nom.) said (refl.) feel (inf.) two possibilities (nom.) exist
'fiór∂ur said that he felt that two possibilities existed'

Thus, Mørck's category of subject is a well-defined category from a morphological point of view, but a poorly defined category from a syntactic point of view. If we want to define the subject purely on morphological grounds, then I am sure that can be justified. We will just have to remember that it is a totally different concept from a subject definition based on syntactic behaviour.

### 2.5 Falk (1995, 1997)

Falk (1995, 1997) bases her investigation of Old Swedish on the fact that nominative arguments of transitive verbs are syntactic subjects (1997:23). About obliqueness she concludes (1997:39):

(7) fornsvenskans obliker fungerade inte som subjekt, utan det oblika ledet ska hela tiden analyseras som objekt. Denna slutsats bygger i hög grad på att det inte föreligger några säkra bevis på oblikens subjektsstatus. Sett ur språkinlärarens perspektiv innebär detta att oblika subjekt är ett märkt alternativ i grammatiken, en egenskap hos grammatiken som är inlärbar endast om positiva evidens föreligger. Jag tycker också att det är ett rimligt antagande att oblika subjekt är ett märkt grammatiskt alternativ, att oblika kasus i första hand signalerar objektskap.

> ’The obliques of Old Swedish did not have the function of a subject, instead the obliques should always be analysed as objects. This conclusion

> is very much based on the fact that there are no conclusive evidence for the subjecthood of the oblique. Seen from the perspective of the language acquirer this means that oblique subjects are a marked alternative in the grammar, a property of the grammar which is only acquired if positive evidence are at hand. I also find it a reasonable assumption that oblique subjects are a marked alternative, that oblique morphological case first and foremost signalizes objecthood.'

It is perhaps not an unreasonable thought that Oblique subjects are a marked alternative, since the languages that have them are probably fewer than those who don't. This argument does pose problems though; first, it seems as if Falk assumes that the language learner has a priori knowledge about what is marked and what is unmarked in the grammar. Second, it seems that Falk assumes that morphosyntactic properties outrank syntactic properties. Third, Falk's assumption that the language learner will only assign subjecthood status to Oblique Experiencers if there is positive evidence is based on the assumption that the language learner assumes a link between morphological case and syntactic functions. Why should the language learner assume such a link a priori?

First of all, as we do not know whether there is such a link, it is as reasonable to imagine that the language learner assumes a link between morphological case and semantic roles. Second, if there is a link between morphological case and syntactic functions, we do not know how that link is constructed, it might be so that that there is a link between nominative and syntactic subjecthood, but it might as well be so that there is a link between nominative and oblique and syntactic subjecthood, meaning that the link is a one-to-one link. Falk gives examples where the nominative is obviously a syntactic object in Old Swedish (1997:40). For these the language learner, if (s)he ever assumed a relation between nominative and subjecthood, must have had to reconsider and must have come to the conclusion that morphological case was not a reliable marking for syntactic functions. It seems more reasonable to me that the language learner used some other strategies to find out about syntactic functions, than the ones Falk is suggesting. The interesting question also arises how the language learner recognized the subjecthood of nominatives of transitive clauses. Nominatives are usually situated in front of the finite verb or immediately following it, but Falk has pointed out that these positions can also be occupied by the object, resulting in the possibility of the language learner interpreting nominatives as objects, since (s)he according to Falk interprets Oblique Experiencers as syntactic objects. I can agree on the fact that the borderline between subjects and objects might have been somewhat more fuzzy in Old Scandinavian than today because of OV word order, but I don't think that morphological case marking was the device used to recognize syntactic functions, anymore than it is used in Modern Icelandic today.

We have no acquisition data on Oblique Experiencers in Old Scandinavian. The fact is that Icelandic and German children do learn (dative) Oblique Experiencers around age 3 (Ute Bohnacker p.c.). Oblique Experiencers are therefore not "marked" for them. At least they are not "marked" in Falk's sense since marked for her seems to mean "not subjects".

Falk's presupposition that the nominative is the subject and every other candidate has to prove itself beyond doubt, has some interesting consequences. The situation arises that the examples where the Oblique Experiencer is situated in front of the finite verb, following the finite verb when something else is topicalized, and in first position in subordinate clauses, cannot be considered evidence for Oblique Experiencers' subjecthood, since objects can occupy these positions. Compare the following examples:

(8) a. mere pina bör **thessom** än hinom
more pain deserve they (obl.) than others
'they deserve more pain than the others' (Falk 1995:206)
b. æn sidhan **hænne** kom innan hugh førbudit, tha vilde hon ...
but since she (obl.) came in mind the prohibition, then wanted she..
'but since she remembered the prohibition, she then wanted ... '
(Falk 1997:36)

(9) a. Thy bör **thet** herra och första til
therefore belongs it lords and princes to
'therefore it belongs to lords and princes'
b. at **thz** wäl behagadhe iudhomen
that it well pleased Jews-the
'that the Jewish were pleased' (Falk 1997:37)

According to Falk's analysis, the nominatives in (9) are subjects, but the Oblique Experiencers in (8) are not subjects. Thereby word order cannot constitute evidence for the subjecthood of Oblique Experiencers. But word order does constitute evidence for the subjecthood of nominatives. It seems to me that a consequence of Falk's approach is a demand for more evidence for the subjecthood of Oblique Experiencers than for the subjecthood of nominatives. If we assume that Oblique Experiencers are subjects, we cannot expect to find *more* evidence for their subjecthood than for nominatives, since we would not expect them to be *more subjects* than nominative subjects. Therefore, because of Falk's presuppositions, her approach simply excludes Oblique subjects a priori. Oblique Experiencers are subjected to more stringent standards than nominative NPs.

### 2.6 Conclusions

It is clear why there has been a debate on the syntactic status of Oblique Experiencers. Different scholars have had different definitions of subject, or different opinions on how a subject should behave.

A disadvantage of the approaches of Faarlund, Kristoffersen, Mørck and Falk, and everyone else who claims that Oblique Experiencers are objects, is that they cannot explain PRO-infinitives, like the ones in examples (10) - (12), formed with verbs which select Oblique Experiencers:

(10) a. fiorgils kva∂st lei∂ast flarvistin Old Icelandic
fiorgils (nom.) said (refl.) be-bored (inf.) there-stay-the (nom.)
'fiorgils said that he was bored by staying there'
b. fiór∂ur kva∂st flykja tvennir kostir til
fiór∂ur (nom.) said (refl.) feel (inf.) two possibilities (nom.) exist
'fiór∂ur said that he felt that two possibilities existed'
(Rögnvaldsson 1996:60)

(11) a. os duger ey ther æptir langa Old Swedish
us helps not that for long
'it does not help us to long for that'
b. huat hiælper idher ther æptir langa
what helps you that for long
'what help is it to you to long for that'
(Falk 1997:25)

(12) a. him burfl to liken well his lif Old/Middle English
he should like well his life
'he should like his life well'
b. good is, quafl Joseph, to dremen of win
good is, said Joseph, to dream of wine
'it is good to dream of wine, said Joseph'
(Seefranz-Montag 1983:133-134)

Although (10)-(12) would standardly be analysed as PRO-infinitives, Faarlund, Kristoffersen, Mørck and Falk are forced to say that these are not PRO-infinitives. Subject-to-object-raising and Subject-to-subject raising will also have to be analysed and labelled differently, both from corresponding constructions in Modern Icelandic and from corresponding constructions of nominatives in Old Scandinavian.

Kristoffersen and Falk argue that in order to assume Oblique subjects in Old Scandinavian, the evidence has to be conclusive. That raises the question what evidence should be regarded as conclusive. Kristoffersen (1991:89) says:

(13) Samanlikna med oblike subjekt i nyislandsk er det snakk om svært få eigenskapar, og eg vil derfor rekne dei utheva ledda [...] som objekt, og ikke subjekt.

> 'Compared with oblique subjects in Modern Icelandic, it is a question of very few properties, and I will therefore count the bold-faced constitutents [...] as objects and not as subjects.'

From Kristoffersen's discussion it is not clear how many subject properties are necessary for the evidence to be conclusive, and it is not at all clear which criteria he uses to decide what amount of evidence is sufficient and what is not.

Falk (1995:203) argues about Modern Icelandic:

(14) Den egenskap som i mitt tycke helt avgörande bevisar det oblika ledets subjektsstatus är att det kan fungera som tankesubjekt i infinitivfraser;

> 'The property which in my opinion is completely decisive on the subjecthood of oblique subjects is the fact that it appears as PRO with infinitives.'

In her 1997 study, though, Falk does not accept the Old Swedish PRO-infinitives as evidence for the subjecthood of Oblique Experiencers, neither does she suggest any other evidence as conclusive. Thereby *conclusive evidence* has become an impossibility in Falk's approach.

It appears that scholars have been applying methods that a priori exclude the result that Oblique Experiencers are syntactic subjects. What has been stipulated as conclusive evidence is very hard to obtain in practice, and when the necessary evidence cannot be found, the conclusion is drawn that Oblique

Experiencers are not subjects, instead of assuming that the methods used don't provide an answer to the question.

Under the approach that Oblique Experiencers are objects, a strange situation arises: conclusive evidence is demanded in order to assume that Oblique Experiencers are analysed as subjects, but conclusive evidence is not needed to analyse them as objects. Kristoffersen, Mørck and Falk have all, implicitly or explicitly, argued along such lines. Some scholars, though, (Seefranz-Montag, Falk) have acknowledged that Oblique Experiencers are not particularly object-like. What can be the reason for this obvious inconsistency in the argumentation? Let's start by considering the following quote from Sundman (1985:104) in (15):

(15) En oblik konstituent kunde "lyftas" ur den inbäddade satsen och placeras i början av matrissatsen, men **den blev inte subjekt** i denna sats; det är därför snarare fråga om en flyttningstransformation än om en relationsförändrande process.[my bolding]

'An oblique constituent could be "raised" from the embedded clause and be placed in the beginning of the matrix clause, **but it did not become subject** in this clause; it is therefore rather a question of a movement than a relation changing process.'

This is an example of the term *subject* being used synonymously with *nominative*. Because the Oblique Experiencer does not change its case from oblique to nominative, it cannot be regarded as a subject, according to Sundman. This could be the root of the problem. Scholars are still defining the syntactic subject as nominative, as if it were a law of nature, and as long as they are, the implicit conclusive evidence for the subjecthood of an NP is nominative case. But of course, if Oblique Experiencers were nominatives, the point I am making would be unnecessary. It is hardly likely that there is a one-to-one relation between morphological case and syntactic functions, since in the languages under discussion, the morphological case system consists of four case forms and there are only two direct syntactic functions. A description of a one-to-one relation between morphological case and syntactic function is perhaps adequate for Modern Mainland Scandinavian and Modern English, but not for Old Scandinavian and Old English, which have a much greater resemblance to Modern Icelandic. The possibility exists that the scholars that have most forcefully rejected the idea that Old Scandinavian had Oblique subjects, are projecting the conditions of Modern Mainland Scandinavian onto Old Scandinavian. Faarlund, Kristoffersen, Mørck and Falk all presupposed that the subject is nominative; their results have also been in accordance with that.

Traditional grammar assumes that the subject is nominative for Old Scandinavian, and that objects are accusative, dative and genitive. In spite of the fact that the category of object divides into three morphological case forms, no-one has ever made the claim that these objects are different kinds of clause constituents. Anyone claiming that the syntactic subject is nominative and nothing but nominative, and that there is a one-to-one relation between morphological case marking and syntactic functions, should also claim, for the sake of consistency, that there exist three kinds of syntactic objects.

## 3 General Methodological Discussion

In *Table 1* above, I illustrate that Oblique Experiencers pass almost all subjecthood tests in Old Scandinavian. That is also what we would expect if these are syntactic subjects, no more and no less. If we do not accept that Old Scandinavian had Oblique subjects despite the fact that they pass the subjecthood tests, the other possible analysis is that they are syntactic objects. Such an analysis entails that we have a group of objects which behave like subjects, with the consequence that the distinction between subjects and objects cannot be maintained. Another option is to say that these subjecthood tests are not subjecthood tests for Old Scandinavian, or to say that the subject did not exist as a syntactic category, such claims are highly disputable. I prefer the analysis that Oblique Experiencers are syntactic subjects in Old Scandinavian, since that analysis does not entail any controversies, apart from putting to question the axiom that the subject is nominative, an axiom which has already been demolished by facts from Modern Icelandic and Modern Faroese.

There are even further methodological problems to be discovered when comparing the above studies, which differ drastically in some respects. My approach, following Rögnvaldsson (1996), is a contrastive, synchronic one (so is also Faarlund's and Kristoffersen's 1994 approach), while the other approaches are simply synchronic. I compare Old Scandinavian with Modern Icelandic, while my opponents are testing their hypotheses against synchronic data. Both approaches can easily be justified. As I see it though, if we take the purely synchronic approach we are faced with one problem: we don't have a starting point. Then we either have to be theoretically tentative, in order to make our own starting point, or we have to have an axiom. The problem with being theoretically tentative is that one can actually reach whatever conclusions one desires, as Mørck does, and the problem with choosing an axiom is that the axiom determines how we treat the data, as in Kristoffersen's and Falk's case. Their axiom is that nominatives are subjects and obliques are objects. Their results are also in accordance with that. Thereby, our approach, and the presuppositions we start with in the beginning decide the results of our investigation.

This can also be said about my approach. Since Oblique Experiencers are analysed as syntactic subjects in Modern Icelandic, I am predestined to reach the conclusion that Oblique Experiencers also were syntactic subjects in Old

Icelandic or Old Scandinavian. I do not deny this. But the advantage of my approach is that it takes the modern language as a starting point. At least we are assuming things for the ancestor language that are "facts" in the modern language, we are not just assuming an axiom. In this case Modern Norwegian cannot be considered a good starting point, even though it is an ascendant of Old Norse, since Modern Norwegian doesn't have Oblique subjects. Perhaps we can draw the conclusion that usually, when we are deciding on an axiom as a starting point for our synchronic analysis of an ancestor language, we choose an axiom which is rooted in the modern language, and thereby the methodological problem discussed here never arises.

The synchronic approach raises even further problems and questions, such as, to what length should we go in being synchronic? Shouldn't anyone who is not willing to use the modern language as a starting point, begin from scratch then? But doing that did not lead Mørck to any reliable results on the subjecthood of Oblique Experiencers! And since in this case an axiom is not a good choice either, aren't we forced to question the subjecthood of nominatives? I seriously doubt that examples of PRO-infinitives can be found for all verbs of Old Scandinavian. Shouldn't we be doubting the subjecthood of the nominative subjects of those verbs because of that? How far should we go in analysing phenomena in the ancestor language as different from the modern language, even though they look alike? What criteria should we use in deciding how far to go before we are driven *ad absurdum*?

More than twenty years have gone since linguists working on Modern Icelandic started revealing that the (at the time) indisputable fact that the subject is nominative could not be maintained (see Andrews 1976). Now we know that there is not a one-to-one relation between morphological case and syntactic function. In my opinion, this should be regarded as a serious threat to the axiom of traditional grammar, that the subject is nominative in Old Scandinavian and Old Germanic.

## 4 Summary and Conclusions

In this article I have compared different methods to determine the (non)subjecthood of Oblique Experiencers in Old Scandinavian.

An examination of the lines of argumentation by other scholars has revealed that the analysis that Oblique Experiencers are syntactic objects is not based on any evidence of them behaving as objects, but merely on their morphological marking as non-nominatives. Obviously, such argumentation is rooted in axioms of traditional grammar that the subject is nominative, an axiom which has already been demolished by facts from Modern Icelandic. The most important thing here is to be aware of what premises one has, since these premises can determine the results of the investigation. My premises are that since Modern Icelandic has Oblique subjects, we can expect that Old Scandinavian also had Oblique subjects, since the structure and case system of Old Scandinavian was a lot like the structure and case system of Modern Icelandic.

Ideology legitimates any methodology, thereby it should be its scientific consequences which ultimately determine our choice of methods.

*Institutionen för nordiska språk*
*Lunds universitet*
*Helgonabacken 14*
*S-223 62 Lund*
*johanna.barddal@nordlund.lu.se*

## Notes

[*] I want to thank the following people for discussions and/or comments: Ute Bohnacker, Joan Maling, Christer Platzack and Eiríkur Rögnvaldsson. I also thank everyone at the Forskarseminar in Lund 15 October 1997 and everyone at the 17th Scandinavian Conference of Linguistics on Fyn 21 Agust 1998, where I presented earlier versions of this work.

[1] Modern German seems to have a case system and a structure similar to Modern Icelandic but nevertheless does not allow Oblique Subjects. Modern German distinguishes itself from Modern Icelandic and Modern Faroese in that almost all constructions with irregular case patterns have been lost from Modern German while Modern Icelandic and Faroese have retained these constructions. In that way the structure of Modern German differs radically from the structure of Modern Icelandic and Faroese.

[2] It should be noted that Faarlund's goal is to compare the Old Norse nominative with the subject of Modern Norwegian, and not necessarily to find out whether Oblique Experiencers are syntactic subjects or not.

[3] In fact it turns out, in a typological survey by Corbett (1991), that languages which have obligatory agreement are forced to make use of some form, usually the most unmarked one, in signalizing agreement (or the lack of it) between the predicate and a non-prototypical controller. Corbett (1991:204) calls this "neutral agreement":

> (i) The range of non-prototypical controllers varies from language to language, ... It may include clauses, infinitive phrases, nominalizations, interjections and other quoted phrases, noun phrases in particular cases (for example, subject noun phrases in an oblique case), dummy elements and certain null elements.

The reason for selecting the most unmarked form as the "neutral agreement form" can be due to the language-specific semantics of genders and persons. The languages which select the neuter form as a neutral agreement form do so because inanimates are evenly distributed over the genders, while animates are denoted by masculine and feminine genders. The neuter form can therefore not contribute to a clash between neutral and human (1991:206). Similarly the "default form" of the verbs taking Oblique Subjects in Icelandic is the only verb form which always coincides in phonological form with the verb form of some other person. There are six verb forms in the present indicative in Icelandic; 1st, 2nd and 3rd person singular, and 1st, 2nd and 3rd person plural. Five phonological/morphological forms are used. That means that one form is used twice in the paradigm. The form used twice is either used to denote 1st and

3rd person singular, or 2nd and 3rd person singular. Applying Corbett's explanation, Icelandic has forced one form to function as a "non-prototypical controller agreement marker", since agreement is obligatory and the form which gets selected is the one and only form used twice in the paradigm. By selecting this "most common" form, the possible clashes get fewer than otherwise.

[4] Verbs selecting Oblique Subjects cannot passivize, but verbs selecting dative objects can of course passivize and then occur with a dative subject. These two are not to be mixed together since it is only the former I am discussing here.

## References

Allen, Cynthia. 1986. Reconsidering the history of *like*. Journal of Linguistics 22:375-409.

Allen, Cynthia L. 1995. Case Marking and Reanalysis. Grammatical Relations from Old to Early Modern English. Oxford: Clarendon Press.

Andrews, Avery. 1976. The VP Complement Analysis in Modern Icelandic. Proceedings of the Sixth Annual Meeting of the North Eastern Linguistic Society: 1-21.

Barðdal, Jóhanna. 1997. Oblique Subjects in Old Scandinavian. Working Papers in Scandinavian Syntax 60:25-50.

Barðdal, Jóhanna. 1998. Argument Structure, Syntactic Structure and Morphological Case of the Impersonal Construction in the History of Scandinavian. Scripta Islandica 49:21-33.

Barnes, Michael. 1986. Subject, Nominative and Oblique Case in Faroese. Scripta Islandica 37:13-46.

Butler, Milton C. 1977. Reanalysis of Object as Subject in Middle English Impersonal Constructions. Glossa 11, 2:155-170.

Corbett, Greville G. 1991. Gender. Cambridge: Cambridge University Press.

Denison, David. 1990. The Old English Impersonals revived. Papers from the 5th International Conference on English Historical Linguistics: Cambridge, 6-9 april 1987. S. Adamson, V. A. Law, N. Vincent and S. Wright (eds.). Amsterdam: John Benjamins.111-140.

Dyvik, Helge J. J. 1986. Grammatikk og empiri. Bergen.

Elmer, Willy. 1981. Diachronic Grammar: The History of Old and Middle English Subjectless Constructions. Linguistische Arbeiten 97. Tübingen: Niemeyer.

Faarlund, Jan T. 1990. Syntactic Change. Toward a Theory of Historical Syntax. Mouton: Berlin.

Falk, Cecilia. 1995. Lexikalt kasus i svenska. Arkiv för nordisk filologi 110:199-226.

Falk, Cecilia. 1997. Fornsvenska upplevarverb. Lund University Press, Lund.

Fischer, Olga C. M. and Frederika C. van der Leek. 1983. The Demise of the Old English Impersonal Construction. Journal of Linguistics 19:337-368.

Fischer, Olga C. M. and Frederika C. van der Leek. 1987. A "Case" for the Old English Impersonal. Explanation and Linguistic Change. W. F. Koopman, F. van der Leek, O. Fischer and R. Eaton (eds.). Amsterdam: John Benjamins. 79-120.

Harris, Alice C. 1973. Psychological Predicates in Middle English. Paper presented at the Annual Meeting of the Linguistic Society of America, San Diego.

Keenan, Edward L. 1976. Towards a universal definition of "Subject". In Subject and Topic, pp. 303-334. Ed. Li, Ch.N., New York: Academic Press, Inc.

Kristoffersen, Kristian E. 1991. Kasus, Semantiske roller og Grammatiske funksjonar i norrønt. Institutt for nordistikk og litteraturvitskap, Oslo: Universitetet i Oslo.

Kristoffersen, Kristian. E. 1994. Passiv i norrønt og nyislandsk - ei sammanlikning. Norsk lingvistisk tidsskrift 12:43-67.

Mørck, Endre. 1992. Subjektets kasus i norrønt og mellomnorsk. Arkiv för nordisk filologi 107:53-99.

Rögnvaldsson, Eiríkur. 1991. Quirky Subjects in Old Icelandic. In Papers from the Twelfth Scandinavian Conference of Linguistics. H. Á. Sigurðsson (ed.) Reykjavík: Institute of Linguistics, University of Iceland. 369-378.

Rögnvaldsson, Eiríkur. 1995. Old Icelandic: A Non-Configurational Language? NOWELE 26:3-29.

Rögnvaldsson, Eiríkur. 1996. Frumlag og fall að fornu. [Subject and Case in Icelandic Middle Ages] Íslenskt mál og almenn málfræði 18:37-69.

Seefranz-Montag, Ariane. v. 1983. Syntaktische Funktionen und Wortstellungsveränderung. Die Entwicklung 'subjektloser' Konstruktionen in einigen Sprachen. Studien zur Theoretischen Linguistik 3. München: Wilhelm Fink Verlag.

Seefranz-Montag, Ariane v. 1984. 'Subjectless' constructions and syntactic change.Historical Syntax. Ed. J. Fisiak. Berlin: Mouton Publishers.

Sigurðsson, Halldór Á. 1989. Verbal Syntax and Case in Icelandic. Doctoral dissertation, University of Lund.

Sigurðsson, Halldór Á. 1992. The Case of Quirky Subjects. Working Papers in Scandinavian Syntax 49:1-26.

Smith, Henry. 1994. 'Dative sickness' in Germanic. Natural Language and Linguistic Theory 12:675-736.

Smith, Henry. 1996. Restrictiveness in Case Theory. Cambridge: Cambridge University Press.

Sundman, Marketta. 1985. Från *Mik angrar* till *Jag ångrar*. Folkmålsstudier 29:85-123.

Traugott, Elizabeth. C. 1982. Syntax. CHEL I, The Beginnings to 1066, pp. 168-289.

Zaenen, Annie, Joan Maling and Höskuldur Thráinsson. 1985. Case and Grammatical Functions: The Icelandic Passive. Natural Language and Linguistic Theory 3:441-483.

# Negation: a cognitive hurdle in interpreting

*Anne Marie Bülow-Møller*

**Abstract**

Compared with a normal positive utterance, negation requires increased cognitive processing for producers and comprehenders of language. This study presents evidence from interpreting. In an experiment with 30 trainees, a large number of negative cues were lost in simultaneous interpretation; these included missing negations, missing inherent negatives, and dislocated negative scope. These units typically made sense internally, but not in context. On the basis of cognitive processing models like the Collaborative Activation-Based Production System (CAPS), it is suggested that under cognitive overload, priority goes to echoic, short term memory and long term frame knowledge, whereas text-based contextual memory is bypassed.

Interpreters do not just translate; they perform a feat consisting of

*a)* *listening* to a stream of source language, and *making sense* of the contribution in terms of speech act and rhetorical relation to context,

*b)* *translating* terms and *mapping* structures unto a target language string, with arguments (actor, goal, etc.) and information structure (given and new information), plus *editing* the unit with time-saving loss of redundancy and superordinate substitution of details, and

*c)* *speaking* the unit aloud while listening to the continued stream.

This multiple activity has been described both in terms of the several 'efforts' involved (Gile 1991) and the 'competition' for what is after all a limited attention span (McWhinney 1997). But there is every reason to study the process also in terms of the *accessibility* of inference and memory in the cognitive load.

Inferences draw on the general knowledge, beliefs, and frames in the comprehender's memory. An inference receives more activation if it draws on a knowledge structure that is already active in working memory: it takes less effort to integrate e.g. a 'cause' in a text chunk that represents a situation that is already 'open' (for an overview of this development since van Dijk and Kintsch (1983), see Graesser and Zwaan 1995). Working memory was investigated already in Kintsch's (1988) "Construction-Integration Model": there is better recall of propositions if they are well integrated in the textbase network of nodes. The more recent "Collaborative Activation-Based Production System (CAPS) Model" holds not only that that working memory is limited, but also that processing at full capacity produces a 'graceful loss' of those elements in working memory that take least part in the processing. The model operates with several levels of activation, rather than all-or-nothing, so that an element must be noted above a certain threshold to take part in the processing at all (Just and Carpenter 1992, Goldman and Varma 1995). This model built on node strength and relative activation can predict recall patterns rather more accurately, and it ties in well with the work on effort and competition in interpreting .

So far, the gracefully lost parts of the input in the documentation from cognitive psychologists have been a matter of mild inaccuracy. In this case, however, recall patterns were flawed by real discrepancies of meaning: there was a notable loss of *negation*, both as sentence polarity and inherent negation (i.e. words that signify an absence).

That negation is in fact cognitively troublesome can be documented from a range of pscholinguistic studies. Thus Clark and Clark (1977) note how in a verification task it takes longer for subjects to look at a *red* spot and judge the claim the *spot is not blue* (which is correct) compared with a positive claim (like *the spot is blue*, even though that is false). Fodor et al. (1975) note how this also goes for inherent negatives like *unmarried*, although not to the same extent. There is a rich literature on the context of negatives to explain this: negatives are used when the speaker addresses an (assumed) hearer expectation that needs correcting (e.g. Ducrot 1972, Givón 1978, 1995, Talmy 1996); the speaker thus holds both the positive and the negative variety in his or her consciousness at the same time. But there is also psychological evidence that sometimes this process goes awry: Brown (1990) demonstrates how with spoken instructions, academically weak adolescents sometimes fail to remember the negation, having fastened on the nouns and verbs, as in *the butler is (not) in the library*.

The present study is reported more fully in Bülow-Møller (forthcoming). It builds on 30 tapes transcribed from trainee interpreters rendering a speech from English into their mother tongue, Danish. The speech concerned the Conservative Heritage; it argued that John Major's government had stolen a number of civil rights from the British people. The subject matter was well known to the interpreters, but the phrase structure was purposefully complicated. In the following, three types of failure are discussed where negation has played a role. None of the examples are one-off occurrences, but – as with all translation – categories are not so clearcut that accurate figures can be given.

### 1. Missing negative.

(The government stole some rights, first over Europe.)

1.1

*then by creating a secret policeforce which wasn't to be distinguished from the Secret Services*

1.2

og så minsandten ved at danne et hemmeligt politi som så igen var en anden organisation end efterretningstjenesten (and then by actually forming a secret police force which was again a different organization from the Secret Service)

1.3

en hemmelig politistyrke som skulle arbejde separat fra den britiske efterretningstjeneste (a secret police force which was to work separately from the British Secret Service)

It is possible that the missing negation was simply not heard, for in ordinary speech the function words get very little prominence compared with semantic content words. This negation was not only contracted (*n't*), it occurred between *was* and *to*, leaving the brief nasal to carry the entire weight of the polarity. In this situation, the interpreter may be in doubt if there was a negative or not, and this may be resolved with recourse to the context: if the argument is clear, one option will make more sense that the other. Here the distinguishing feature of the new *secret policeforce* is its secrecy; its working methods are what makes the force newsworthy, as an instance of invasion of rights on a par with the existing law-unto-themselves Secret Services. For interpreters who missed this negation, it seems that they *bypassed* the evidence of the context.

## 2. Missing inherent negative

2.1

*Out went the right of citizens to remain silent during their trial. In came the rights of the police to bug the premises of doctors, lawyers and journalists.*

2.2

skal folk ha lov at være tavse under retssagen (are people to be allowed to be silent during the trial)

2.3

loven gik ud på at folk sku være stille under retssagen ( a law to the effect that people are to be silent during the trial)

2.4

*So, where is that 'Nation at peace with itself' that John Major promised in 1990? Far from being at peace with itself, Britain has not been so divided before.*

2.5

selvom øh Storbritannien er meget fredfyldt nu så er det osse delt op (although Britain is very peaceful now it is also divided)

2.6

fra at hvile i sig selv har Storbritannien ikke været så skilt som nu (from being at peace with itself , Britain has never been so separated as it is now.)

In both these examples (2.1 and 2.4) the inherent negative represents a contrast: *far from being x, y,* and *out went x, in came y.* If the first word goes missing we have structures like *from (existing) x we move to y* and *did x go?* It cannot be a case of not hearing in both cases – certainly *out* was under contrast stress each time the speech was given. It can, however, be a case of dismissing a problematic adverb as not part of the sentence argument structure (i.e. fastening on nouns and verbs, as in Brown's data.) Another explanation is in the information structure: the marked-theme slot is normally filled with given, accepted information. With *out went the right to remain silent*, the information is genuinely new in context; normally, new evidence in this position should be positive, as there is as yet no expectation to contradict.

In 2.4, *at peace with itself* is not actually framed as a surface negative. It would have been possible to read the rhetorical question as *Where is that nation that used to be at peace with itself?*, except that it does not make sense as a promise by John Major. Again, with access to the context it should have been possible to realize that *where is x?* presupposes that *x is not here.*

## 3. Scope displaced

3.1
*The decision to join EMU is not a monetary issue - it is not a mathematical calculation whether we give up our independence and sovereignty for a good deal.*

3.2
Beslutningen om at blive medlem af EMU'en er ikke en matematisk beregning hvor vi skal gi op vores øhm uafhængighed og suverænitet (the decision to join EMU is not a mathematical calculation where we are to give up our independence and sovereignty).

3.3
At følge ØMU'en, det er ikke en monetær politik, det er ikke et spørgsmål om at gi vores ø:hm uafhængighed og suverænitet øh -- (to follow EMU, that is not a monetary policy, it is not a question of giving our independence and sovereignty uh --)

It is arguable that these dislocated negations reflect the interpreter's beliefs, i.e. the general knowledge that so far no 'giving up of independence and sovereignty' has occurred. In the difficulty of having a choice of candidates to map the negation scope unto, echoic memory is flooded, and since the *monetary issue* is not quite an 'open' file, the chunk has to be processed from long-term beliefs. The crucial structural feature about the passage is its rhetorical parallel: the meaning depends on the referential overlap between *monetary issue* and *mathematical calculation,* which gives the second sentence the status of 'elaboration' on the first. If the cumulative textbase was available for the comprehender, it would be clear that these two constitute the 'same' negation, not a negative that is balanced by another negative, or followed by a negative that has *giving up* in its scope (as with the translations that end up as *it is not a question of us giving up Britain's sovereignty*).

It would seem, then, that in some cases it is the negation itself that does not make it over the 'threshold', so that it is not processed both in memory and in reproduction; in other cases it is presumably the presence of the negative that takes up processing capacity so that other elements receive less attention and the scope of the negation is displaced. In both cases, brief reflection would reveal that the translation is nonsense in context; but under the severe time pressure that simultaneous interpreting represents, this is the loop that trainees have not mastered. The distribution suggests a hierarchy, with processing effort being invested primarily in short-term, echoic memory items and long-term knowledge, so that the single sentence contains no information that the interpreter knows to be wrong; the middle ground, the cumulative text base, seems to be bypassed.

*Copenhagen Business School, Department of English*
*DK - 2000 Frederiksberg*
*e-mail: amb.eng@cbs.dk*

## References

Brown, Gillian. 1990. Text comprehension: Missing the cues. Journal of Applied Linguistics 6. 37-47.

Bülow-Møller, Anne Marie. Forthcoming. Existential problems: how irrealis is processed in simultaneous interpretation.

Clark, Herbert H and Eve V Clark. 1977. Psychology and Language. An introduction to psycholinguistics. New York: Harcourt Brace Jovanovich

Danks, Joseph H, Gregory M Shreve, Stephen B Fountain and Michael K McBeath (eds). 1997. Cognitive Processes in Translation and interpreting. Thousand Oaks, CA: Sage.

van Dijk, Teun A. and Kintsch, Walter. 1983. Strategies of Discourse Comprehension. London: Academic Press.

Fodor, JD, Fodor, JA and Garrett, MF.1975. The psychological unreality of semantic representation. Linguistic Inquiry 6:4. 515-31.

Gile, Daniel. 1991. The processing capacity issue in conference interpretation. Babel 37.1, 15-27.

Gile, Daniel. 1997. Conference interpreting as a cognitive management problem. In Danks et al (eds), 196-214.

Givón, T. 1978. Negation in language: Pragmatics, function, ontology. Syntax and Semantics vol 9: Semantics, ed. Cole. New York: Academic Press.

Givón, T. 1995. Functionalism and Grammar. Amsterdam: Benjamins.

Goldman, Susan R and Sashank Varma. 1995. CAPping the Construction-Integration Model of Discourse. In Weaver et al (eds), 337-58.

Graesser, Arthur C. and Rolf A. Zwaan 1995. Inference generation and the construction of situation models. In Weaver et al (eds), 117-39.

Just, Marcel A and Patricia A Carpenter. 1992. A capacity theory of comprehension: Individual differences in working memory. Psychological Review 99:1. 122-49.

Kintsch, Walter. 1988. The role of knowledge in discourse comprehension: A construction-integration model. Psychological Review 95. 163-82.

Kintsch, Walter. 1993. Information accretion and reduction in text processing. Discourse Processes 16:1-2. 193-202.

MacWhinney, Brian 1997. Simultaneous interpretation and the competition model. In Danks et al (eds), 215-32.

Talmy, Leonard. 1996. The windowing of attention in language. Grammatical Constructions. Their form and meaning, in eds. Shibitani and Thompson. Oxford: Clarendon Press.

Weaver, Charles A, Suzanne Mannes and Charles R Fletcher (eds) 1995. Discourse Comprehension. Essays in Honor of Walter Kintsch. Hillsdale, NJ: Erlbaum.

# Inflection classes, Norwegian verbs and Cognitive Grammar

*Hans-Olav Enger*

## Abstract

The importance of product-orientation in inflection is advocated, and the notion of 'inflection class' is discussed. The notion of similarity is claimed to be important in inflectional morphology.

## Source-oriented versus product-oriented generalisations

We shall begin by considering the distinction between source-oriented and product-oriented generalisations. Source-oriented generalisations are statements to the effect that an element X undergoes a process Y to produce a new form Z. Neither X nor Z needs to be well-defined, but Y is. Product-oriented generalisations deal with Z, neither X nor Y is well-defined. An example of a source-oriented generalisation: 'To get present-tense forms in Norwegian, add /-r/ to the stem.' (Stem=X, adding /-r/ = process Y, present-tense forms=new form Z). The corresponding product-oriented generalisation would be: 'Present-tense forms in Norwegian usually end in /-r/.' (See also Nesset 1998: 23-6 for a good introduction to the declarative/procedural distinction, and compare the discussion of declarative versus procedural approaches in phonology and syntax.)

Source-oriented models have so far been predominant in theorising on inflectional morphology. My claim is that if we rely on source-orientation alone, along with a 'standard' conception of inflection classes, generalisations will be lost. I am not saying that source-orientation should be dispensed with;

but that it is not enough. The problems with source-orientation are intertwined with the discussion of inflection classes.

## Introduction

An inflection class is a 'group of words or stems that inflect in the same or a similar fashion' (Matthews 1991:129f). This definition entails some problems:
*Inflection classes are posited in order to bring out what is common in the inflectional behaviour of two or more lexemes. But sometimes, what actually happens is the opposite; i.e., the classes obscure what is common.
*Sameness and similarity are different concepts.
*How many is a group?

## 1 Inflection classes can be unhelpful

Inflection classes help us capture similarities between forms that are said to be members of the same class. Unfortunately, inflection classes are not **always** helpful when it comes to capturing similarities. Once two lexemes have been assigned to different classes, the classes do not help us capture similarities between forms of these two lexemes.

In Swedish, for example, there is a set of past tense forms ending in /u:g/: *tog* /$^{1}$tu:g/ 'took', *drog* /$^{1}$dru:g/ 'pulled', *log* /$^{1}$lu:g/ 'laughed', *dog* /$^{1}$du:g/ 'died', *slog* /$^{1}$slu:g/ 'hit'. The corresponding stems (which are identical to the infinitive forms) are not similar: *ta* /$^{1}$ta:/ 'take', *dra* /$^{1}$dra:/ 'pull', *le* /$^{1}$le:/ 'smile', *dö* /$^{1}$dø:/ 'die', *slå* /$^{1}$slo:/ 'hit'. Therefore, the vowel changes one must posit to get from the infinitive stems to the past tense forms are dissimilar.

It is easy to generalise over *ta* and *dra*, because they have similar stems and similar past tense forms. The problem is how to capture the similarity with *log*, *dog* and *slog*. Within a strictly stem-based model of morphology combined with atomic inflection classes, there is apparently no way to account for this similarity, and Kiefer (1970:124) explicitly labels the past tense forms *log*, *dog* and *slog* suppletive: Thus, they emerge as completely isolated and unique. Consequently, a generalisation is missed, namely that these past tense forms are similar to each other and to those of *ta* and *dra*. It is an admirable aspect of Kiefer's analysis that it is carried out in such detail that these limitations of the model are brought out.

The generalisation we wish to bring out is that certain past tense forms are similar. This is a product-oriented generalisation. The problem is that the only way in which the framework adopted by Kiefer can relate inflected forms is by assigning these forms to the same class, and you cannot assign forms to the same class unless their base forms (i.e. the infinitive forms, in this case) or the processes they undergo are similar.

Admittedly, one possible line of defence for a strictly source-oriented model would be to label the similarity between /$^1$lu:g/, /$^1$du:g/ and /$^1$slu:g/ spurious. However, this line of defence seems untenable, for several reasons. First and foremost, any generalisation can be spurious, but a generalisation should not be dismissed just because one particular model is not able to capture it. Secondly, in the case at hand, there is no independent justification for saying that the similarity between *tog* and *drog* is more interesting than that between *tog* and *log*. In fact, there is evidence that product-oriented generalisations are useful. In the case of Swedish past tense forms in /u:g/, the schema has also been productive: We also find /$^{1}$stu:g/ (of the verb *stå* 'stand'), in which case the /g/ is an innovation. Generally, the fact that a generalisation is productive, indicates that the generalisation is a valid one. So the product-oriented generalisation is probably valid.

There are several cases in which product-oriented generalisations are useful. Consider the Norwegian Bokmål verbs *si* /$^{1}$si:/ 'say', *legge* /lege/ 'lay' and *be* /$^{1}$be:/ 'beg, pray'. The infinitive forms do not have much in common, but the past tense forms /$^{1}$sa:/, /$^{1}$la:/ and /$^{1}$ba:/ are similar. In the supine, only /$^{1}$sakt/ and /$^{1}$lakt/ resemble each other. Or take *si* /$^{1}$si:/ 'say' and *bringe* /$^{2}$briNe/ 'bring'. Their supines are similar - /$^{1}$sakt/ and /$^{1}$brakt/ - even if their infinitive, present and past tense forms are not. Another example is *falle* /$^{1}$fale/ 'fall' and and *smelle* /$^{2}$smele/ 'make a bang'. The infinitives do not look similar, but their past tense forms are, cf respectively /$^{1}$falt/ and /$^{1}$smalt/.

When we establish inflection classes, we tend to assume that if two verbs are inflectionally similar in one form, they are so in all. But this is not always the case, as the examples above show. Some verbs are similar in one form, e.g. *le* /$^{1}$le:/ 'laugh' and *stå* /$^{1}$sto:/ 'stand' (which resemble each other only in the past tense). Some resemble each other in two forms (e.g. *bringe-tvinge*, which are similar in the infinitive and present). Some resemble each other in three forms (*se-be*, which are quite similar everywhere except the past). And some - in fact most - resemble each other in all the four forms normally listed in Scandinavian grammars.

Those verbs that look similar in all four forms, do not create any problems for the source-oriented models. And those are the majority. It is hardly accidental that the majority fits the standard pattern. But there remains an interesting minority which creates problems for any strictly source-oriented classification; which does not mean only Kiefer (1970); in fact, it holds for most Norwegian grammarians.

For instance, Venås (1967:110) says that since the past tense vowel of *falle* is *a*, the verb belongs to class III, i.e. the same class as *smelle*. But Venås does not normally group verbs together only on account of similarity in one form. For instance, Venås (1967:211) refuses to group together classes IV and V, on the grounds that their supine vowels are different, even though these classes are similar in both the infinitive, the present and the past. In the case of *falle* and *smelle*, by contrast, he groups together verbs which differ formally in both the supine and the infinitive. This is inconsistent. Venås's classification is usually source-oriented, but in the case of *falle*, he tries to capture a product-oriented generalisation.

The upshot is that product-oriented generalisations are needed in order to capture similarities between classes. Such generalisations have to some extent been left unexpressed within source-oriented models.

This may be one reason why grammarians tend to disagree when it comes to the number of inflection classes and their members in any given language (according to Plank 1991:27),. Some product-oriented generalisations are simply too obvious. Yet, if one tries to capture them, the result may be inconsistency. (The problems with **establishing** inflection classes are discussed at more length by Bittner 1996 and Enger 1998, from somewhat different points of view.)

## 2 Inflection classes are not diacritics

We now turn to the psycholinguistic status of inflection classes. Within a realist context, i.e. where the linguist's constructs are claimed to have something to do with what goes on in the minds of speakers, this is not unproblematic. This is easily overlooked. For example, Kiefer (1970) takes the traditional concept 'strong' in Swedish verb morphology to be psychologically valid without any discussion. He also assumes that speakers utter some strong past tense form because that particular verb is stored in their brain with the diacritic [+strong]. Consider the following:

> except for the features [+/-Strong] and [+/-Pres=Sup], no other morphological feature is needed in order to account for the ablaut of the Swedish strong verbs. This would also correspond to what the native speaker of Swedish would know about a 'new verb', provided he gets some more information about this verb. If he were told that it is a strong verb and its stem vowel is *u, y* or *a* he would without hesitation 'generate' the right forms.
>
> (Kiefer 1970:133)

Here, the role of language acquisition is virtually ignored, for it is unclear how a native speaker can learn the stem of a verb and the fact that it is strong without being exposed to any of its forms. It seems reasonable to demand that at least one actual form should be listed in the mental lexicon, and that is fairly uncontroversial these days. Yet parallels to Kiefer's liberal use of features are found in more recent literature. For example, Durand (1990:123) uses the feature [+D] to generate the correct past tense form of the English verb *drink*. Durand says the function of this and similar 'arbitrary symbols is merely to trigger the correct allomorphy rules.' But we may wonder about the effect of this move. Traditionally, one would simply say that the past tense of *drink* is *drank*. By using the feature [+D], Durand implies that a generalisation has been captured, but the feature [+D] covers only this one verb, and there is no independent evidence for [D], i.e. the feature is established *ad hoc*. To be fair;

on this point Durand's textbook simply echoes the approach of many other works, but the idea is unacceptable. In a similar vein, Katamba (1993:102) says that '[i]f we know the class that a verb belongs to, we can predict how it is going to be affected by ablaut in the past tense and past participle.' But 'predict' is hardly the appropriate word here. It is on the basis of the past tense and participle forms that we can tell what class a verb belongs to, not vice versa. To claim that a past tense form is predictable on the basis of the class the verb belongs to, or that [+D] explains a past tense form such as *drank*, is, to put it harshly, equivalent to claiming that facts are predicted from our description. There is no independent evidence for such features, and the whole idea seems untenable.

Within Cognitive Grammar (cf Langacker 1987 and elsewhere) such descriptions are outlawed (due to the Content Requirement). And it is difficult to agree when classes are turned into separate entities in the mental lexicon, with the function of 'triggers' and a separate existence besides being a representation of similarity. It is not because a verb is marked [+ strong] in the lexicon that it has a monosyllabic past tense form; the relation is the other way round.

To summarise so far: Saying that a given word (lexeme) belongs to a particular inflection class, is not much different from saying e.g. that its past tense form is reasonably entrenched in a certain form (cf Langacker 1987:422). There are reasons for preferring the more 'concrete' way of phrasing it, as far as psychological plausibility is concerned. In short, the idea that classes are represented as diacritics triggering rules, seems unappealing.

## 3 Similarity

An inflection class is defined as a group of words that inflect in the same or a similar fashion. This definition is more intriguing than it may seem. Firstly, there is nothing logically discrete about the notion 'be like, be similar' (Givón 1986:90, Wurzel 1990:87). Secondly, the notion is actually absent from many morphological models. For instance, in the Word-and-Paradigm model as we know it from Matthews (1991), there is no way of capturing similarity. Within this model, there are only two options, viz identity and difference. Either forms are due to the same rule, or they are due to different rules. And yet Matthews's definition of inflection class reads 'same or similar'.

One may wonder why scholars have focused so much on sameness, and so little on similarity. It is not obvious that similarity should be equated with sameness. Similarity is a much broader notion than sameness. Presumably, the reason why sameness (identity) has been focused is that this concept is easily formalised; similarity is not (cf Wurzel 1990, who makes the argument explicit). In my view, this argument is like letting the tail - formalisation - wag the theoretical dog.

Now, there are two options: either we change the definition of inflection class so as to get rid of the fuzzy notion of similarity, or we take the notion of similarity seriously, and incorporate it in our morphological model. It is the latter course that is advocated here. Since similarity is a fuzzy notion, accepting similarity presupposes that we accept a 'non-classical' model of categorisation. This has been disputed on methodological grounds, but it must be acceptable.

Similarity may even be a key concept in morphology. For instance, a definition of morphology could be that it is the study of formal and semantic similarities that go in parallel. Since similarity is not a discrete notion, this would explain why there are no firm boundaries of morphology, and the frequently noted fact that it is so difficult to delimit the field. The idea that morphological relations are gradual is also supported by studies in language acquisition; for example, this is probably what Peters & Menn (1993:771f) mean by suggesting that the morpheme is a continuum.

The notion of inflection class, then, relies on the notion of inflectional similarity. Since 'similar' is not ordinarily a discrete notion, we would not expect inflectional similarity to be so, either, and consequently, nor should inflection classes be discrete. Now, linguists may define the technical term 'inflectionally similar' or 'inflection class' so as to turn it into a discrete notion. This is seldom done, however.

Another problem concerning inflection classes may be that the very metaphor of a 'class' leads us to expect entities that are sharply distinguished from one another on all points. We tend to think of classes as separate entities. But inflection classes are not entirely separate. Classes can share properties, i.e., they can be similar on some points. Some classes are closer to one another than others, as the historical evidence shows in Norwegian.

Usually, a verb that shows inflectional similarity with another in one form, does so in all. Normally, then, there is a clear correlation. But the fact that such correlations are not necessary, shows a drawback as regards the metaphor of classes, namely that it leads us to expect this correlation as a necessary one. This is what gives us problems in the case of *tog, stog*.

At the same time, there is an obvious tendency for a strong present tense form to correlate with a strong past tense form. Given prototype theory, this is to be expected, according to Givón (1986): Between absolutely discrete and distinct categories on the one hand and absolute non-discreteness and fuzziness on the other, is the expectation that 'properties tend to coincide statistically/probabilistically, but do not always coincide absolutely' (Givón 1986:79). The Swedish example we started out with is an illustration: Usually verbs have the same inflectional properties in both the infinitive, the present tense, the past tense and the supine -- but not always.

## Conclusions

I do not mean to 'reject' the notion of inflection class as such. I have meant to focus on some problems with it, and to advocate the importance of product-orientation. A product-oriented approach to morphology is also more in line with declarative approaches towards language, which have become more common both in phonology and in syntax. The importance of similarity in syntax is also advocated by Langacker (1997). It may also be relevant in word-formation (cf Enger 1997).

Generally, there must be something between sameness (identity) and unrelatedness (difference), namely similarity (or resemblance or relatedness). And there must be some way of expressing similarity.

*Institutionen för Nordiska Språk*
*Stockholms universitet*
*S-106 91 Stockholm*
*Hans_Olav.Enger@nordiska.su.se*

## References

Bittner, Andreas 1996. Starke 'schwache' Verben -- schwache 'starke' Verben. Tübingen: Stauffenburg.

Bybee, Joan L 1985. Morphology. Amsterdam and Philadelphia: John Benjamins.

Durand, Jacques. 1990. Generative and Non-linear phonology. London: Longman.

Enger, Hans-Olav 1997. On subtractive back-formation in Scandinavian nouns. ZAS Papers in Linguistics, vol 8, September 1997, 44-56.

Enger, Hans-Olav 1998. The classification of strong verbs in Norwegian with special reference to the Oslo dialect: A study in inflectional morphology. Oslo: Universitetsforlaget.

Givón, Talmy 1986. Prototypes: Between Plato and Wittgenstein. Craig, C (ed): Noun classes and categorization. Amsterdam and Philadelphia: John Benjamins.

Katamba, Francis 1993. Morphology. Basingstoke: Macmillan.

Kiefer, Ference 1970. Swedish morphology. Stockholm: Skriptor.

Langacker, Ronald W 1987. Foundations of Cognitive Grammar, vol 1. Stanford: Stanford University Press.

Langacker, Ronald W 1997. Constituency, dependency, and conceptual grouping. Cognitive Linguistics 8. 1-32.

Nesset, Tore 1998. Russian Conjugation Revisited. Oslo: Novus.

Matthews, P H 1991. Morphology. Second edition. Cambridge: Cambridge University Press.

Peters, Ann M & Lise Menn 1993. False starts and filler syllables: ways to learn grammatical morphemes. Language 69. 742-777.

Plank, Frans 1991. Of abundance and scantiness in inflection. Plank, Frans (ed): Paradigms,1-39. Berlin: Mouton de Gruyter.

Venås, Kjell 1967. Sterke verb i norske målføre. Oslo: Universitetsforlaget.

Wurzel, Wolfgang. Ullrich. 1990. Gedanken zu Suppletion und Natürlichkeit. Zeitschrift für Phonetik, Sprachwissenschaft und Universalienforschung 43. 86-91.

# Conditionals in Institutional Settings

*Tarja Salmi-Tolonen*

**Abstract**

Conditionals do not always take the prototypical form of a conditional in English but there are several constructions, which nevertheless express condition; furthermore the position of the conditional in the sentence is relevant to the discourse function of the conditional. This paper explores the variety of forms and investigates their use in English legislative discourse in national and supranational contexts. The differences between the conditionals in the two types of texts are found to be both qualitative and quantitative. The discourse functions of the preposed conditionals seem to differ from those positioned sentence finally.

## 1. Introduction

The purpose of the paper is to ascertain similarities and differences in the function of conditionals depending on clause order, and to explore the variety of surface structure expressions and their use in different positions in institutional settings. It is of particular interest to see how forms function in texts in the public domain and institutionally grounded texts, in this case legislative texts. There can even be practical implications gained from this kind of study, if it can contribute to our understanding of the organization of information in the legal settings from which the empirical data have been drawn. In this study conditionals are scrutinized in two legal contexts: the directive and rights-conferring context (British national) and the cooperative and harmonizing context (supranational, the European Union). (1)

Crystal and Davy (1969: 181) comment that in legislative writing 'every action or requirement ... is hedged around with and depends upon a set of conditions which must be satisfied before anything at all can happen'. Hart

(1961: 35), a legal theorist, argues that: 'There is no law prohibiting murder: there is only a law directing officials to apply certain sanctions in certain circumstances to those who do murder.' Hart goes on to say that the 'if-clause' is not directed to ordinary citizens, but to officials and orders them to apply certain sanctions if certain conditions are satisfied. 'All genuine laws, on this view, are conditional orders to officials to apply sanctions. They are all of the form, "If anything of a kind X is done or omitted or happens, then apply sanction of a kind Y."'(ibid.). It comes natural therefore that conditionals are of particular interest to legal linguistics.

## 2. Conditionals

Most of the studies on conditionals agree that in describing and interpreting conditionals the following domains are essential: knowledge, causality, time, and modality. Below I will discuss some of these approaches which are of particular interest or relevance to the description of legislative conditionals.

The conditional is constructed of the *protasis* or the prerequisite and its logical consequent, the *apodosis*, in other words 'If X (protasis) then Y (apodosis)' as based on our real-world knowledge. Expressing the logical relationship between the two is, however, not the only function.

In English the overt marker of conditionality is 'if', the only other possibility being the subject verb inversion. The apodosis in English can be marked with 'then', although not necessarily. There are, however, several other constructions in English which are not overtly marked as conditionals, but which through the context can be interpreted as conditionals. Below I shall focus on the constructions appearing in the present corpora.

Sweetser (1990: 113 ff) suggests a tripartite classification for conditionals and calls them *content, epistemic, and speech act conditionals*. By the content conditionals she means the 'real-world' type of conditionals. 'If Paris is the capital of France then two is an even number' is not a well-formed conditional because there is no connection between the two clauses. The fulfilment of the protasis, likely or unlikely, is a sufficient condition for the fulfilment of the apodosis. In the following example the causal link between the two clauses is not at the content level, but at the epistemic level. 'If she's divorced, then she's been married' (op.cit.: 116). In the case of speech act conditionals the performance of the speech act represented in the apodosis is conditional on the fulfilment of the state described in the protasis as in 'If I may say so, that's a crazy idea'. How then are the conditionals in statutory language related to these three categories?

Legal conditionals are, of course, of a particular kind, less in a formal sense, except in the way that the linguistic forms used are limited, but more in the sense that they are removed from our real-world knowledge, existing only in the 'text-world' at hand. They are not necessarily common sensical and form a logical sequence. The protasis expresses not only a sufficient condition but a necessary condition and the apodosis often expresses enablement rather than cause. The content is only true because it is the will of the legislator (cf. examples 1 and 2). The reader can draw some conclusions from the context of the protasis globally and locally and therefore I contend that the statutory conditionals fall into Sweetser's category of epistemic conditionals.

(1) If the defendant is not domiciled in a Contracting State, the jurisdiction of the courts of each Contracting State shall, subject to the provisions of Article 16, be determined by the law of that state. (EC)

(2) If the terms of a contract under which a woman is employed at an establishment in Great Britain do not include (directly or by reference to a

collective agreement or otherwise) an equality clause, they shall be deemed to include one. (NL)

The above examples represent the prototypical cases, the *casuistic* forms. Even the conditionals with an if-marker are not always as uncomplicated as this. There are cases where the protasis and apodosis are not adjacent to each other and, of course, the apodosis may precede the protasis.

Attempts have also been made to categorize conditionals according to the verb forms used in conditional expressions. Dancygier (1993) suggests that the interpretation of conditionals is made more specific with respect to the use of verb forms. However, Dancygier limits her discussion to conditionals, which are overtly marked and identified on two accounts: the presence of 'if' in the protasis and the syntactic frame, namely a complex sentence. Dancygier's classification of English conditionals involves tense and modality and she claims that conditionals fall into two major categories: 'predictive' and 'non-predictive'. The following are examples of predictive conditionals (op.cit.: 405):

(3) If it rains, the match will be cancelled.

(4) If it rained, the match would be cancelled.

(5) If it had rained, the match would have been cancelled.

Legislative conditionals are often of a form similar to the sentence in example (3) (cf. examples 1 and 2), and, as far as tense goes, of this type only. In the protasis there is the present tense, and the modal auxiliary is 'shall'. This, although used as modal verb of obligation, still encompasses futurity. The other legislative modal is 'may' expressing possibility. Furthermore, it is debatable whether legal conditionals are predictive or non-predictive. Dancygier (op.cit.: 417) claims that non-predictive conditionals refer to the time they indicate and are therefore not back-shifted. By backshifting she means that the time reference intended by the speaker is systematically later than the time referred to by the verb form in its prototypical (non-conditional) uses (op.cit.: 406).

Categorization of conditionals according to the time perspective presents problems. First of all, there are different perspectives of time and the time perspective in law is not exclusively linear, which, however, would appear to be Dancygier's basic assumption in her categorization, and which is generally the case in everyday language use. I would suggest that legislative conditionals be both predictive and non-predictive. For the interpretation of legislative conditionals two points of time are essential: the point of time when the legislation is enacted and the point of time it is applied (cf. also Wilson and Sperber 1997). A legal conditional carries both predictive and non-predictive meaning because it states the legal consequences should the premise expressed in the protasis come true at some point of time in the future, and it also functions as a link to the past at the time of application, because at that point the premise has been fulfilled.

Dancygier (1993: 425) claims further that non-predictive conditionals lack iconicity, in other words the sequence of clauses does not have to be iconic of the sequence of events and that non-predictive conditionals are essentially non-causal. At the propositional content level they lack the element of causality. However, Dancygier admits that causality may have some role in the processing of a non-predictive construction, especially one that is inferential. In the interpretation of legislative conditionals for a legal norm the cause-effect relationship is essential.

The prototypical form of a conditional clause even in legislative writing is a complex sentence starting with a dependent clause, which expresses the condition, the protasis. A legal norm is to be generally applicable and this is expressed by the present tense. The apodosis expressing the legal consequence is always unconditional and therefore it must be in the indicative or imperative.

I suggest that there are basically four types of conditional constructions in English which might be called *casuistic* (cf. Clark 1974) as above in examples 1 and 2 representing a causal relationship between the two parts of the expression; *relative* starting with a relative pronoun 'who', 'whoever', 'where', or 'wherever' and being more generalized in meaning, providing background knowledge or presenting the case; the third type is the *apodictic* type expressing in fact a command accompanied with a sanction and taking a form of a command; the fourth type being what I have called the *concessive* type and which is marked by such expressions as 'except', 'apart from', 'save as' (see Table 1). The concessive conditionals express an exception to the norm, which has already been regulated before an exception to the rule is allowed.

In general, it has been considered an elementary rule of legal draftsmanship that the same thing must be expressed in the same words; the convergent usage of words and forms is essential. The data studied here presents evidence of the use of a variety of forms, as the examples show. The morphologically and semantically different conditional expressions make statutory language formally diverse. Rather than prove that the modern draftsmen have failed in their task, this may indicate that there is a conflict between the conventional requirement of consistency and the pressure of cognitive factors in legal writing.

## 3. Conditionals in discourse

Haiman (1978) argued that the antecedents of conditionals function pragmatically as given or old information in discourse and therefore are topics. Akatsuka (1986: 342 ff), on the other hand, claims that conditionals are not given and that Haiman's analyses do not shed light on the meaning of conditionals. Criticism against Haimanís views has also been presented by Ford and Thompson (1986) and Ramsay (1987). The reason for such contradictory views may be, apart from the fact that the notions have been given different meanings by different investigators, that none of these studies seem to take into account that conditionals can also function simultaneously at the discourse, syntactic, semantic and pragmatic levels. Noh (1998) gives an interesting relevance-theoretic account of conditionals and in particular metarepresentative uses in conditionals, which he argues will enable us to make the correct interpretations. These views, which take us towards a more pragmatic analysis of conditionals, although beyond the scope of this paper, are certainly worth exploring.

### 3.1. Initial conditional clauses

It has been suggested that the scope of the initial, or preposed adverbials have a broader-scoped anaphoric textual grounding and are thematically associated to the preceding discourse (e.g. Givon 1990 and Thompson 1985). Preposed adverbial clauses may be viewed as coherence bridges at major thematic junctures. Their bridging capacity is further aided by their cataphoric semantic connections to the main clause. Furthermore, a sentence initial conditional can provide shared knowledge for the following material in the case in which the conditional offers a contrast to something which has gone before as in the following examples from the studied data.

Type 1. X. If X, then Y.

(6) 1. - (1) A person discriminates against a woman in any circumstances relevant for the purposes of any provision of this Act if ...

(a) ...

(b) ...

If a person treats or would treat a man differently according to the man's marital status, his treatment of a woman is for the purposes of subsection (1)(a) to be compared to his treatment of a man having the like marital status. (NL)

Type 2. X. If not X, then Y.

> (7) 10.-1. For the purposes of this Part and section 1 of the Equal Pay Act 1970 ("the relevant purposes"), employment is to be regarded as being at an establishment in Great Britain unless the employee does his work wholly or mainly outside Great Britain.
>
> (3) In the case of employment on board a ship registered at a port of registry in Great Britain ... the ship shall for the relevant purposes be deemed to be the establishment.
>
> (4) Where work is not done at an establishment it shall be treated for the relevant purposes as done from which it is done or ... at the establishment with which it has the closest connection. (NL)

Example 4 beginning with the locative relative 'where' represents what we have called a relative type of a conditional clause. These are said to have a more generalized meaning than the *casuistic* type of conditionals in normative texts (see eg. Clark 1974).

Type 3. X. If alternative Y, then Z.

> (8) 49 (1) If an organisation to which section 12 applies comprises a body the membership of which is wholly or mainly elected, nothing in section 12 shall render unlawful provision which ensures that a minimum number of persons of one sex are members of the body - (NL)
>
> (9) 6 (3) Except in relation to discrimination falling within section 4, subsections (1) and (2) do not apply to employment for the purposes of a private household, or ... (NL)
>
> (10) If the formal investigation is one required by the Secretary of State the Commission shall ... (NL)
>
> (11) If the formal investigation is not one required by the Secretary of State, the Commission shall ... (NL)

The initial conditionals in the data serve as framework for the following clause, either in direct reaction to something in the preceding context or exploring options relevant to the situation expressed in the preceding context.

Exemplifying discourse-related phenomena poses problems since space seldom allows texts to be presented in their entirety. In example 12 the if-clause in Subsection 4 reaches across sentence boundaries to all previous subsections in Section 59, the linking clauses have been underlined in the text extract. At the same time the if-clause refers cataphorically to its main clause and the rest of the sentence. A sentence-initial conditional can sometimes have a purely cataphoric function as exemplified by the except-clause in example 12.

(12) 59. -(1) For the purposes of a formal investigation the Commission, by a notice in the prescribed form served on him in the prescribed manner, -

(a) may require any person to furnish such written information as may be described in the notice, and many specify the time at which, and the manner and form in which, the information is to be furnished;

(b) may require any person to attend at such time and place as specified in the notice and give oral information about, and produce all documents in his possession or control relating to any matter specified in the notice.

**(2) Except as provided by section 69, a notice shall be served under subsection (1) only where -

(a) service of the notice was authorised by an order made by or on behalf of the Secretary of State, or

(b) the terms of reference of the investigation state that the Commission believe that a person named in them may have done or may be doing acts of all or any of the following descriptions -

(i) unlawful discriminatory acts,

(ii) contraventions of sections 37,

(iii) contraventions of sections 38, 39 or 40, and

(iv) acts in breach of term modified or included by virtue of an equality clause,

and confine the investigation to those acts.

(3) A notice under subsection (1) shall not require a person -

(a) to give information, or produce any documents, which he could not be compelled to give in evidence, or produce, in civil proceedings before the High Court or the Court of Session, or

(b) to attend at any place unless the necessary expenses of his journey to and from that place are paid or tendered to him.

*(4) If a person fails to comply with a notice served on him under subsection (1) or the Commission has reasonable cause to believe that he intends not to comply with it, the Commission may apply to a county court for an order requiring him to comply with it or with such directions for the purpose as may be contained in the order; and section 84 (penalty for neglecting witness summons) of the County Courts Act 1959 shall apply to failure without reasonable excuse to comply with any such order as it applies in the cases there provided.

### 3.2. Final conditional clauses

As opposed to initial conditional clauses final or postposed adverbial clauses are only related to their main clause and thus have a very localized scope.

Type 4. X if Y

(13) (1) A person discriminates against a woman in any circumstances relevant for the purposes of any provision of this Act if on the ground of her sex he treats her less favourably than he treats or would treat a man, or ... (NL)

In definitions at the beginning of the Act.

(14) This section does not apply if the discrimination only concerns employment which the employer could lawfully refuse to offer the woman. (NL)

Ford and Thompson claim that when a conditional clause occurs within a nominalization, an infinitive, or a relative clause, there is a tendency for it to occur in final position. And they offer as one factor working against a conditional clause appearing in an initial position the tendency for an 'interesting' subject to be introduced in a nondependent rather than a dependent clause. In legislative text the the motivation seems rather to be the degree of importance of the information and also the fact that the condition is limited to the main clause of the sentence and does not reach beyond sentence boundaries.

## 4. Conclusion

Various linguistic structures are used for expressing the same function even in statutes drafted within the same context. The differences between the texts produced in the two different contexts studied are primarily quantitative but to some extent also qualitative. Considering conditional clauses in terms of frequency and position in the sentence we find that the EC corpus displays a greater number of initial clauses as opposed to the final conditional clauses whereas in the NL corpus the numbers are equal. This seems to indicate a difference in text strategies in the corpora.

The discourse function and the position of the conditional seem to be intertwined even in legislative discourse. Sentence-initial conditionals seem to perform a strategy, which brings background information for the subsequent propositions in the discourse. The scope of the initially positioned adverbials is more global in the discourse and therefore they are not as closely connected with their main clause as the adverbials in the final position. Preposed adverbial clauses also function anaphorically at the discourse level as coherence devices. However, they also have a cataphoric function through their syntactic and semantic connections to the main clause, whereas final or postposed adverbial clauses are only related to their main clause and thus have a very localized scope.

The paper gives only a partial and incomplete answer to the function of conditionals in legislative writing. There are several aspects of legislative conditionals, which deserve to be scrutinized more thoroughly. These include the time perspective, hypotheticality and the cognitive constraints behind the choice of the type of conditional and its positioning in the text.

European Community law (EC) corpus consists of articles concerning equality in the Treaty of Rome, and two directives on the same subject; the national corpus, NL, consists of the Equal Pay Act 1970 and the Sex Discrimination Act 1975.

*FIN- 20014 University of Turku, Finland*
*Tel. +358 2 333 8760, Fax +358 333 8730,*
*E-mail tasato@utu.fi*

## References

Clark, C. 1974. Law In Hayes, J (ed.) Old Testament Form Criticism. London.

Crystal, D. and D. Davy. 1969. Investigating English Style. London.

Dancygier, B. 1993. Interpreting conditionals: Time, knowledge and causation. Journal of Pragmatics 19, 403-434.

Ford, C.E. and S.A. Thompson. 1986. Conditionals in discourse: a text-based study from English. In Traugott, E.C. et al. (eds) On Conditionals. Cambridge.

Givón, T. 1990. Syntax. A Functional-Typological Introduction. Volume II. Amsterdam.

Haiman, J. 1978. Conditionals are Topics. Language vol. 54-3,564-589.

Hart, H.L.A. 1961. The Concept of Law. Oxford.

Noh, E.-J. 1998. A Relevance-theoretic Account of Metarepresentative Uses in Conditionals. In Rouchota, R. and A.H. Jucker (eds.) Current Issues in Relevance Theory. Amsterdam.

Sweetser, E. 1990. From etymology to pragmatics. Cambridge.

Thompson, S.A. 1985. Grammar and written discourse: Initial vs final purpose clauses in written English discourse. Text 5:1-2. 55-84.

Wilson, D. and D. Sperber 1997. Pragmatics and Time. In Carston, R. and S. Uchida (eds.) Relevance Theory; Applications and Implications. Amsterdam.

**Table 1. Distribution of different types of conditionals in initial, medial, and final positions.**

| | I | | M | | F | | total f |
|---|---|---|---|---|---|---|---|
| | EC | NL | EC | NL | EC | NL | |
| if | 25 | | 16 | | 18 | | 59 |
| | | 17 | | 25 | | 31 | 73 |
| where | 20 | | 4 | | 3 | | 27 |
| | | 29 | | 4 | | 10 | 43 |
| when | | | | | | | |
| | | 1 | | | | | 1 |
| whenever | 2 | | | | | | 2 |
| | | 1 | | | | | 1 |
| whether | | | | | 2 | | 2 |
| | | | | 3 | | | 3 |
| unless | 1 | | 1 | | 3 | | 5 |
| | | | | 4 | | 14 | 18 |
| except | 1 | | | | | | 1 |
| | | 5 | | 3 | | | 8 |
| in the case | | | | | 2 | | 2 |
| | | 2 | | 4 | | 2 | 8 |
| notwithstanding | 6 | | | | | | 6 |
| | | | | 1 | | | 1 |
| ~if but only if | | | | 2 | | | 2 |
| ~if-less | | 2 | | | | | 2 |
| *save as | 3 | | | | | | 3 |
| *provided that | | | 1 | | 2 | | 3 |
| *in the absence of | | | 1 | | | | 1 |
| *should | 2 | | | | | | 2 |
| *apart from | 1 | | | | | | 1 |
| total | 61 | | 23 | | 30 | | 114 |
| | | 57 | | 46 | | 57 | 160 |
| | | | | | | | |
| ~occurs only in the NL corpus | | | | | | | |
| *occurs only in the EC corpus | | | | | | | |

# Different competences? A comparison of the productive and receptive language of a Danish agrammatic patient

*Elisabeth Engberg-Pedersen and Lise Randrup Jensen*

### Abstract

A case study of a Danish patient with so-called agrammatic aphasia shows that her productive language is not unconstrained by rules or principles, even if these do not correspond to standard Danish. In spontaneous speech production, her choice of initial argument tends to be semantically restricted by her choice of predicate. Passive sentences are never produced, either in spontaneous speech or in a passive elicitation task. The patient's comprehension is investigated with a set of tests designed to illuminate the role of semantic factors and argument order in her comprehension of active as well as passive sentences.

## Introduction

In aphasiology a certain type of aphasic speech is traditionally termed agrammatic. Compared with the standard language, the speech of agrammatic aphasics appears unsystematic. Characteristic features are: reduced phrase length, deviating or absent morphology, lack of function words, and reduced complexity of sentence structures (see Berndt 1991 for a review). Some aphasic patients with agrammatic speech also have difficulties understanding aspects of meaning which is signalled by grammatical structure and morphology, i.e. they appear to be agrammatic in both comprehension and production (Caramazza & Zurif 1976). However, research has shown that there are patients with so-called agrammatic speech who have relatively normal comprehension, just as there are patients who comprehend agrammatically without showing the symptoms of expressive agrammatism (Goodglass 1968). A general conclusion on research with agrammatic aphasics has been that there is no general set of symptoms shared by all patients with this diagnosis. Rather, there are individual differences in the precise nature of the expressive as well as the comprehension problems (Berndt 1987).

The traditional theories of agrammatism have been focused on the problems with morphology and function words, and word order has been said to be unaffected (Goodglass & Menn 1985). Two studies by Saffran et al. (1980) and Schwarz et al. (1980) have, however, demonstrated considerable problems with word order in the comprehension and production of sentences in a group of agrammatic patients. In their study of sentence production, Saffran et al. (1980) showed that the agrammatic aphasics appeared to have preserved a superficial ability to conform to the prototypical nominal-verb-nominal pattern of simple English declarative sentences, but they were unable to order the nominals around the verb so that the position of the nominal consistently reflected the thematic role in which its referent was depicted in the stimulus materials. The pattern of ordering nominals around the verb was not, however, random. The patients in the study seemed to resort to an animacy-based strategy in sentence production such that an animate noun would be placed before the verb even when the referent of the animate noun was depicted in another role than agent in the picture materials.

In this study we present data from the production and comprehension of sentences by a Danish woman, KE, who had a stroke in 1990 at the age of 35 and whose speech shows the characteristic problems of agrammatic patients. Through a case study of KE's productive and receptive language we would like to address the general question of whether her language is indeed agrammatic in the sense of being unconstrained by rules or principles. We shall especially focus on this question in relation to word order and the issues raised by Saffran et al.'s demonstration of the influence of animacy. To the extent that KE's expressive language appears to be rule-governed, we will also pose the question of whether KE's language comprehension follows the same principles as her productive language.

## KE's productive language: spontaneous speech

KE has many of the symptoms typical of agrammatic aphasia. Her productive language is dominated by uninflected lexemes as can be seen in example (1) from KE's retelling *Frog, Where are you?* (Mayer 1969):

(1) *dreng og hund og en få [mange forsøg på at udtale* frø*]*
boy and dog and a [få] [many attempts at pronouncing *frø* 'frog']

*frø (3) aften (P) (rømmer sig) aften fordi øh måne (2)*
frog (3) evening (P) (clears her throat) evening because uh moon (2)

*og øh nu tal nul bog? (...) frå elsker nat (P) til dug og hyggeligt*
and uh now speak zero book? (...) [frå] love night (P) to dew and cozy

In her retelling of the Fro-story, KE uses some function words, but not consistently. In (1) the indefinite article *en* only appears before the noun *frø* 'frog' and not before any of the other nouns. The preposition *til* 'to' seems to be a marker of the genitive, prepositional objects, and all sorts of locative relations. All lexemes in (1) are uninflected, but KE does use inflected lexemes such as verbs in the present tense (*sover* 'sleep', *elsker* 'love') and nouns inflected for number (*huse* 'houses'). But again, the inflections are not used consistently, not all verbs are inflected and the plural *huse* is used about the interior of one house.

But KE is a creative language user. When she is in doubt whether she is allowed to look at the pictures while telling the story, she asks *nu tal nul bog?* 'now speak zero book' with an unorthodox negation and a topic-comment structure reminiscent of topic-prominent languages (Li & Thompson 1976), but not found in standard Danish. Moreover, she has a very good grasp of the story. Her version of the Frog-story includes all three elements that Berman and Slobin use to characterize story-tellers' ability to produce a thematically coherent narrative with a global structure (1994:46ff.). Moreover, when KE had finished telling the story, we gave her the 24 pictures in random order and asked her to reassemble the book, and she did so unhesitatingly, even noticing that by mistake we had left out one picture.

In spite of her obvious word-finding problems, KE has a rich vocabulary using, for instance, such an infrequent, almost obsolete Danish adjective as *bly* 'bashful, shy'[1], and she tells the story with many details and such background information as that frogs love nights because they enjoy the dew.

Danish is a verb-second language. In narrative the constituent before the verb is generally the subject, an adverbial or a pronominal object. In KE's clauses with only one argument that argument precedes the predicate, but occasionally she uses inversion in a verb-first construction:[2]

(2) a. *dreng og hund sover nat*
boy and dog sleep night
b. *falde hund* (inversion)
fall dog
c. *dreng nå* (speaker followed by direct quotation)
boy well

In clauses with two arguments we see the patterns in (3). The initial argument is the argument which is highest on the animacy scale[3] or which is perceptually more salient as the figure or theme in relation to the ground or location.

(3) (Figure +) predicate (+ *til* 'to') + ground

a. *og bikube falde til øh jord*
and beehive fall to uh ground

(Agent/experiencer +) predicate (+ *til* 'to') + recipient/patient

b. *og dreng k- kalde til (P) øø frø*
and boy c- call to (P) [øø] frog

(Possessor +) predicate + possessee

c. *træ har (2) øh hul?*
tree has (2) uh hole?

The initial argument may be left out when it is also the initial argument of the preceding predicate (see (4b) below).

KE uses the lexical difference between *bange* 'afraid' and *skræmme* 'frighten' to vary the choice of initial argument in the examples in (4) thereby changing the perspective:

(4) a. *dreng måske skræmme til ugle* (*dreng* = agent)
boy maybe frighten to owl

b. *og igen bange til dreng til til din bør- børn* (omitted argument = experiencer)
and again afraid to boy to to your chil- children

In (4b) *ugle* 'owl' is the initial argument of the preceding clause and the sentence probably means 'The owl is afraid of the boy for its children's sake.'

Besides (2b) with inversion and no other constituent before the predicate, the word order of all the preceding examples is acceptable in standard Danish. But being a verb-second language, standard Danish varies the initial constituent not only by means of voice, but also in active sentences by placing, for instance, an adverbial or the object initially. Fronting of particularly adverbials with subsequent VS-order is seens even in young children's retelling of the Frog-story. By contrast, KE in almost all cases sticks to a semantically predictable choice of initial argument or she leaves out this argument altogether, a possibility that is not acceptable in uncoordinated sentences in standard Danish.

In the Frog-story KE uses one predicate with a variation in the semantic status of the initial argument. The verb *kigge* 'look' is used with the agent/experiencer in initial position, but in (5) *gardiner* 'curtains' is not the agent/experiencer:

(5) *og øh gardiner åbent / gardiner kigge*
and uh curtains open+neuter.sg. / curtains look

Example (5) occurs in a context where KE is searching for a verb equivalent to the English verb *to peep*. She explains the verb by describing a situation where you peep through a curtain. The semantic relation of *gardiner* 'curtains' to *åbent* 'open' and *kigge* 'look' in (5) is somewhat doubtful, but the agent, i.e. some generalized notion like *man* 'people, one', is clearly suppressed and *kigge* 'look' is used in its base form as many of KE's verbs. The constructions in (5) are reminiscent of standard-Danish intransitive constructions with a patient subject such as *kartoflerne koger* 'the potatoes are boiling' vs. the transitive *jeg koger kartoflerne* 'I am boiling the potatoes', both in the active voice. The construction *gardiner kigge* is not acceptable in standard Danish, however.

Except for (5) KE's choice of initial argument is semantically predictable from her choice of predicate in such relatively free productive speech as narration on the basis of a wordless picture-book: in sentences with two arguments the more active argument, the one higher in animacy, or the more figure-like argument is placed initially, i.e. the first position signals an agent, experiencer, possessor, or figure relation to the verb. Moreover, KE never uses passive marking, which according to Functional Grammar (Dik 1989:219) means that she does not have a grammatical subject.

## KE's productive language: elicitation of active and passive constructions

In an attempt to elicit passive constructions we presented KE with 10 still-photos of two people interacting, e.g. a girl kissing a boy or a man measuring a girl. In 5 of the photos the active person was placed to the right, in 5 to the left. Earlier research has demonstrated that passive sentences can be elicited by means of verbal instructions varying the focus of attention (Turner & Rommetveit 1967, MacWhinney 1977). In our study each photo was used twice, once with the instruction *Fortæl hvad pigen/drengen gør* 'Tell us what the girl/boy is doing' intended to elicit active sentences, and once with the instruction *Fortæl hvad der sker med drengen/pigen* 'Tell us what is happening to the boy/girl' intended to elicit passive sentences. The photos and instructions were presented in random order.

In 18 out of 20 cases KE started talking about the individual that was mentioned in the elicitation question. One example can be seen in (6) with topicalization of the boy:

(6) Exp.: *Fortæl hvad der sker med drengen.*
'Tell us what is happening to the boy'
KE: *dreng sid øh bord*
'boy sit uh table'

KE's responses in this test demonstrate that she is capable of choosing a perspective dependent on what her interlocutor says. But only in one case did she use a construction with the patient preceding the verb:

(7) Target sentence: *pigen bliver kildet af drengen* 'the girl is tickled by the boy'
KE: *pige kilder* [makes tickling movements with her hands]
'girl tickle
... *fordi bror til hånd til krop til kilder og dame le*
... because brother to hand to body to tickle and lady laugh'

Here *pige* 'girl' expresses the patient argument. But KE did not use passive marking in any case. With passive elicitations she started talking about the passive individual's physical position (e.g. the boy is sitting on the table) or the individual's state of mind as judged by his or her appearance. In 9 out of 10 responses she went on to describe what the agent was doing as in (7). Moreover, in 5 out of these 9 cases she used *fordi* 'because' to link the initial description of the passive individual's state with the description of the active individual's action as in (7). By contrast, in only 3 out of 10 answers to active elicitation she also described what the passive individual was doing or his or her state of mind. It is not possible to know whether her tendency to mention what the active individual is doing even in responses to passive elicitation is due to a wish to describe the transitive action behind a passive construction as a response to the elicitation *Fortæl hvad der sker med ...* 'Tell us what is happening to ...', or whether it is evidence of a tendency always to describe an active participant's action even with passive elicitation (cf. Verfaille & Daems 1996 for experiments testing the priority of the agent in visual event perception).

Summarizing the results from the study of KE's language production, we found that her choice of initial argument tends to be semantically restricted depending on her choice of predicate. She does not spontaneously produce sentences with passive marking, but she is capable of choosing a perspective by alternating her choice of predicate, and she also responds differently to the same picture depending on her interlocutor's choice of perspective.

We were therefore interested in finding out whether the semantic restrictions on the choice of initial argument in KE's productive language also apply to her comprehension. First, we wanted to find out whether she understands passive constructions with the argument lower in animacy in initial position. Second, we were interested in finding out how fine-grained her semantic distinctions are. That is, is her choice of initial argument determined by the argument's status as denoting a human being, the prototypical agent in control of an action, or does she realize that inanimate forces may act on human beings and thus may be more appropriate as initial arguments with some verbs?

To investigate these two questions we designed an anomaly judgment task involving three types of arguments, human beings, forces, and things. Our prediction was that she would prefer the human argument as the initial argument and that she would therefore tend to reject active sentences with a force subject such as *Lynet rammer drengen* 'The lightning hits the boy' and passive sentences with a thing subject such as *Mælken bliver drukket af kvinden* 'The milk is drunk by the woman'.

## KE's language comprehension: the anomaly judgment task

The material consisted of 8 sentences within each of 5 groups varying the choice of initial argument and each occurring in an active and a passive variant.

The resulting 80 sentences were presented in random order and KE was asked to tick off either *OK* or *vrøvl* 'nonsense' next to each sentence. The sentence types and the results appear in Fig. 1.[4] As it is appears, KE can be correct both when judging a sentence to be OK (categories 1 and 2) and when judging it to be nonsense (categories 3, 4, and 5).

Correct sentences according to standard Danish:

1a. force acting on human being, active - 3/8:
*Lynet rammer drengen* 'The lightning hits the boy'

1p. force acting on human being, passive - 7/8:
*Drengen bliver ramt af lynet* 'The boy is hit by the lightning'

2a. human being acting on thing, active - 6/8:
*Kvinden drikker mælken* 'The woman drinks the milk'

2p. human being acting on thing, passive - 2/8:
*Mælken bliver drukket af kvinden* 'The milk is drunk by the woman'

Anomalous sentences according to standard Danish:

3a. human being acting on force, active - 5-6/8:
*Drengen rammer lynet* 'The boy hits the lightning'

3p. human being acting on force, passive - 3/8:
*Lynet bliver ramt af drengen* 'The lightning is hit by the boy'

4a. thing acting on human being, active - 8/8:
*Koppen flytter barnet* 'The cup moves the child'

4p. thing acting on human being, passive - 8/8:
*Barnet bliver flyttet af koppen* 'The child is moved by the cup'

5a. lexically impossible (except as metaphor), active - 8/8:
*Lampen åbner stenen* 'The lamp opens the stone'

5p. lexically impossible (except as metaphor), passive - 8/8:
*Stenen bliver åbnet af lampen* 'The stone is opened by the lamp'

Fig. 1. The types of sentences occurring in the anomaly judgment task. 3/8 means that KE had 3 correct answers out of 8.

KE correctly rejected all sentences of the categories 4a, 4p, 5a, and 5p. In the other cases we find 1. categories where she tends to be correct such as 1p with 7 correct answers out of 8; 2. categories where she tends to be wrong such as 2p with 2 correct answers out of 8; and 3. categories where her performance seems to be random such as 1a with 3 correct answers out of 8. Her unhesitating rejection of the sentences in 5a and 5p demonstrates that she is quite capable of deciding when something is nonsense on purely lexical grounds. The rest of the results can be summarized as shown in Fig. 2.

*Sentences involving human being and force (1a, 1p, 3a, 3p)*
active: human subject (3a, 5-6/8 nonsense): small tendency to be correct
force subject (1a, 3/8 OK): tendency to be random
passive: human subject (1p, 7/8 OK): tendency to be correct
force subject (3p, 3/8 nonsense): tendency to be random
*Sentences involving human being and thing (2a, 2p, 4a, 4p)*
active: human subject (2a, 6/8 OK): tendency to be correct
thing subject (4a, 8/8 nonsense): correct
passive: human subject (4p, 8/8 nonsense): correct
thing subject (2p, 2/8 OK): tendency to be wrong

Fig. 2. Summary of the results from the anomaly judgment task.

The anomaly judgment task demonstrates that KE does not have the strategy that the initial argument should be human: in 3 active sentences in category 1a and 5 passive sentences in category 3p she accepts a force argument as the initial argument. She also accepts two passive sentences with an initial argument denoting a thing in category 2p and correctly rejects passive sentences of category 4p with an initial human argument, i.e. sentences such as *Barnet bliver flyttet af koppen* 'The child is moved by the cup'. So we can reject the hypothesis that KE always prefers an initial human argument.

The results also demonstrate that she is generally better at judging the correct distribution of arguments in sentences involving human beings and things than in sentences involving human beings and forces. In sentences involving human beings and things she is fully correct in 2 out of 4 categories and furthermore tends to be correct in 1 category. In sentences involving human beings and forces she has trouble with the sentences where the force argument comes first and the human argument is placed after the predicate no matter whether the sentences are active or passive (categories 1a and 3p). These sentences were of course designed to split up the two usually converging properties of being active and being human on two different arguments. The results confirm that in this language comprehension test KE tends to have trouble with initial force arguments. Notice, however, that this does not mean that she simply accepts sentences with initial human arguments. In 3a (i.e. sentences such as *Drengen rammer lynet* 'The boy hits the lightning') she has a small tendency to be correct in rejecting the sentences in spite of the initial human argument.

In the categories involving a human being and a thing KE correctly accepts *Kvinden drikker mælken* "The woman drinks the milk' and rejects *Barnet bliver flyttet af koppen*'The child is moved by the cup' and she correctly rejects *Koppen flytter drengen* 'The cup moves the boy', but she is incapable of judging the status of *Mælken drikkes af kvinden* 'The milk is drunk by the woman' which she characterizes as nonsense. This means that she can correctly judge the difference between active and passive sentences when the initial argument denotes a human being, but not when the initial argument denotes a thing where she incorrectly rejects 6 out of 8 passive sentences (category 2p).

We see the same tendency to be better at judging the difference between active and passive sentences with initial human arguments in the sentences involving a human being and a force. Here she tends to be correct or has a small tendency to be correct when the initial argument denotes a human being, but tends to be wrong when it denotes a force.

To sum up,

- when the initial argument denotes a force, KE tends to be wrong;
- when the initial argument denotes a thing, she judges active sentences correctly, but tends to be wrong with passive sentences;
- when the initial argument denotes a human being, she distinguishes active from passive sentences, most convincingly when the second argument denotes a thing, but also when it denotes a force.

Put differently, the results indicate that KE is capable of recognizing and correctly interpreting passive morphology when the initial argument is human. It is as if she expects the initial argument to be human, and when this expectation is met, she has processing capacity to judge the difference between active and passive morphology and even to correctly reject sentences with a human subject.

Shortly after her stroke, KE was taught some Danish Sign Language, and she spontaneously gestures a lot. Moreover, while taking the test she pointed out herself that she visualized the scenes behind the sentences in the anomaly test. In sign languages many verb signs are made iconically with the hands moving outwards from or inwards towards the signer's body. For instance, in the verb meaning 'give' in Danish Sign Language the hand moves outwards from the signer's body in a sentence meaning 'I give to somebody' and inwards in a sentence meaning 'somebody gives to me'. But the sign form with the hand moving inwards towards the body may also be used in a sentence meaning 'she or he received something', i.e. the signer's own body is so to speak used to represent another individual in a construction that is reminiscent of a passive sentence in that it topicalizes the recipient. It is of course only possible to topicalize arguments in this way when they are of a kind that can be personified, i.e. when they can be represented by the signer (Engberg-Pedersen 1996).

We do not wish to claim that KE has taken over the system from Danish Sign Language. Rather the same visual system of an egocentric space where the speaker or signer can identify with the indvidual spoken of seems to lie behind both Danish Sign Language and KE's comprehension of the sentences in the anomaly test. When the human argument is topicalized, she is capable of distinguishing correctly active marking denoting a movement outwards from the (active) body from passive marking denoting a movement inwards towards the (passive) body.[5]

## Constructing sentences with word cards

In a further attempt to test KE's productive rules, we asked her to sort cards with single words. She got three cards in random order and was asked to make a sentence that made sense. All the sentences were active sentences that had also occurred in the anomaly judgment task and they were again presented in random order. In this test KE could manipulate the words and try out different orders.

In 9 out of the 16 sentences she only made one version which in all cases was the correct one. But in one case, *kvinden drikker mælken* 'the woman drinks the milk', she spent more than three minutes in making up her mind and she expressed uncertainty about the result although it was in fact the correct sentence. Only in one case did she end up with a wrong sentence: *pigen blænder lyset* 'the girl blinds the light'. She did not spend more time on testing different solutions in constructing the sentences involving a force argument than the sentences involving a thing argument. Moreover, she constructed *manden serverer kaffen* 'the man serves the coffee' and *pigen spiser æblet* 'the girl eats the apple' although she rejected both as nonsense in the anomaly judgment task. In all 8 sentences with a thing argument, she started out by placing the card with the human argument in initial position and did not change this choice subsequently. In these sentences, she was primarily uncertain about the right order of the verb and the thing argument in second and third position.In sentences with a force argument, by contrast, she tried out a solution with the human argment as the initial argument on 8 occasions, only once ending up with this solution, however.

The results here contrast with the results from the anomaly judgment task where KE tended to judge sentences such as *Lynet rammer drengen* 'The lightning hits the boy' as nonsense. It is somewhat uncertain whether the wordcard sorting task can properly be regarded as a production task. Besides reading the different suggestions aloud herself, KE also asked one of the experimenters to read them to her. Moreover, she was told that the wordcards could be combined to sentences that made sense. In the anomaly judgment task every new sentence might be right or wrong. Thus the processing demands differ in the two cases and this may explain the difference in KE's handling of especially sentences with a force argument. However, even in the word sorting task she seems governed by a preference for an initial human argument.

In conclusion, KE's productive language is not random, so if grammatical means 'constrained by rules or principles', KE's speech is not agrammatic. Nor is her comprehension entirely unsystematic, but the differing results from the various types of language production and comprehension tasks raise the question of whether KE has different competences for different types of tasks or whether the differences can be explained in terms of processing demands.

There are obvious differences in the principles necessary to characterize her performance in production and in different types of comprehension tasks, but the principles seem to revolve around the prominence of semantic factors like humanness in determining or interpreting the order of nominal arguments relative to the verb. Although KE never produces sentences with passive marking, she is capable of interpreting active and passive marking in language comprehension as assessed in the anomaly judgment task under the condition that the sentence begins with a human argument and the second argument denotes a non-human entity, preferably a thing. In sentence construction with word cards KE performed better than on anomaly judgments, correctly solving most of the sentences which she had previously rejected. However, her strategy when trying out different solutions to the force-as-subject sentences showed a clear preference for placing the human argument in the initial position.

The fact that KE's decoding of grammatical marking is facilitated by humanness in the initial argument of the sentence can be illuminated by a parallel in Danish Sign Language, both possibly being based on visual perception and the conceptualization of egocentric space.

*Department of General and Applied Linguistics*
*University of Copenhagen*
*Njalsgade 80*
*DK-2300 Copenhagen S.*
*Denmark*
*e-mail: eep@cphling.dk, lrj@cphling.dk*

## Notes

1. The word does not occur in Maegaard and Ruus (1981), a list of the 5000 most frequent words in a corpus of 250.000 words from Danish novels.

2. In segmenting KE's discourse into sentences, we have used both prosodic cues, her use of *og* 'and' as a marker of sentence start, her tendency to use *til* 'to' to link arguments following the predicate to the predicate and to link nouns with nouns as well as her dominant theme-rheme structure. The segmentation did not present significantly more problems than segmentation of spontaneous discourse in standard Danish.

3. Comrie (1989:Chapter 9) demonstrates that animacy as a linguistic notion is an inherent property of nominals. Linguistic animacy is independent of the relation between the nominal and its predicate, but has been found to control a number of phenomena including case and number marking in some languages. Linguistic animacy correlates with several extra-linguistic parameters, only one of which is the absolute distinction between animate and inanimate.

4. The sentences in category 4a can be understood as OVS sentences in Danish, i.e. they are not nonsensical. The OVS order is highly infrequent with two nouns, however, (rather than one or two pronouns) and out of context most people will interpret them as SVO sentences and thus as nonsensical as did KE in all 8 cases.

5. Comrie (1989:192f.) shows that some languages, e.g. Southern Tiwa and Nahuatl, obligatorily have passive constructions when the patient argument outranks the agent argument in animacy. In Nahuatl the equivalent of *The potato burnt the girl* must be in the passive voice, *The girl was burnt by the potato*, demonstrating the close link between the argument highest in animacy and the point of view.

## References

Berman, Ruth A., & Dan Isaac Slobin. 1994. *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hilsdale, New Jersey: Lawrence Erlbaum.

Berndt, R. S. 1987. Symptom co-occurrence and dissociation in the interpretation of agrammatism. *The cognitive neuropsychology of language* ed. by M. Coltheart, G. Sartori & R. Job, 221-233. London: Erlbaum.

Berndt, R.S. 1991. Sentence processing in aphasia. *Acquired Aphasia* ed. by M. T. Sarno, 223-270. New York : Academic Press.

Caramazza, A., & Zurif, E. B. 1976. Dissociation of algorithmic and heuristic processes in language comprehension: evidence from aphasia. *Brain and Language* 3. 572-582.

Comrie, Bernard. 1989. *Language Universals and Linguistic Typology*. 2nd ed. Oxford: Blackwell.

Dik, Simon. Simon C. 1989. *The Theory of Functional Grammar. Part I: The Structure of the Clause*. Dordrecht: Foris.

Engberg-Pedersen, Elisabeth. 1996. Iconic motivations in conflict: Language-specific structure and influence from the medium. *Content, Expression, and Structure: Studies in Danish Functional Grammar* ed. by E. Engberg-Pedersen, M. Fortescue, P. Harder, L. Heltoft & L. Falster Jakobsen, 39-64 Amsterdam: John Benjamins.

Goodglass, H. 1968. Studies in the grammar of aphasics. *Developments in Applied Psycholinguistic Research* ed. by S. Rosenberg & J.H. Koplin, 177-208. New York: MacMillan Co.

Goodglass, H., & Menn, L. 1985. Is agrammatism a unitary phenomenon? *Agrammatism* ed. by M.-L. Kean, 1-26. New York: Academic Press.

Li, Charles, & Sandra A. Thompson. 1976. Subject and Topic: A new typology of language. *Subject and Topic* ed. by V. N. Li, 457-489.

MacWhinney, Brian. 1977. Starting points. *Language* 53. 152-167.

Maegaard, Bente, & Hanne Ruus. 1981. *Hyppige ord i danske romaner*. København: Gyldendal.

Mayer, Mercer. 1969. *Frog, Where Are You?* New York: Dial Books for Young Readers.

Saffran, E. M., M. F. Schwartz, & O. S. M. Marin. 1980. The word order problem in agrammatism: II. Production. *Brain and Language* 7. 263-280.

Schwartz, M. F., E. M. Saffran, & O. S. M. Marin. 1980. The word order problem in agrammatism: I. Comprehension. *Brain and Language* 7. 249-262.

Turner, Elizabeth Ann, & Ragnar Rommetveit. 1967. Experimental manipulation of the production of active and passive voice in children. *Language and Speech* 10(3). 169-180.

Verfaillie, Karl, & Anja Daems. 1996. The priority of the agent in visual event perception: On the cognitive basis of grammatical agent-patient asymmetries. *Cognitive Linguistics* 7-2. 131-147.

# Was läßt sich im Deutschen *aufhören* und *beenden*? Aspekt, Aktionsart und nominale Komplementation intervallabschlußdenotierender Phasenverben

*Volkmar Engerer*

**Abstract**

The notion of 'phase verb' is not very well established, at least in German grammar. If phase verbs are considered at all, they are constrained to infinitive constructions. It is argued that this syntactic limitation is not valid for German and that nominal complements are also sensitive to phase semantics. This is supported, among other things, by the fact that only a very few phase verbs take sentential, i.e. infinitive, complements, whereas nearly all allow or demand nominals. Furthermore, it can be seen that the aspectual properties of German phase verbs ('Aktionsarten') differ significantly from their English counterparts; this results, for example, in the impossibility of expressing the termination of an accomplishment in German, which is quite natural in English.

'Phasenverb' (im folgenden 'PV') ist als Terminus in der deutschen Grammatikschreibung nicht gut eingeführt. Wenige Grammatiken erwähnen diese Verbkategorie überhaupt und gängige linguistische Nachschlagewerke ignorieren sie meistens ganz. Ausnahmen sind hier die didaktische, in der DDR entstandene Grammatik von Helbig/Buscha (1984), die - so scheint es - den Begriff aus der russischen Tradition übernommen hat (vgl. Mulisch et al. (1985) zum Russischen), sowie die neue, dreibändige IDS-Grammatik (Zifonun et al. 1997). Eine Durchsicht der beiden Grammatiken zeigt, daß die Kategorie PV auf Infinitiv-Einbettungen (*zu arbeiten beginnen* (PV) vs. *die Arbeit beginnen*, kein PV) und in Zifonun et al. (1997: 1390ff) zusätzlich noch auf (infinitivische) Konstruktionen mit impersonaler Belegung des Matrixsubjekts eingeschränkt wird (*Die Lage beginnt sich zu entspannen* (PV) vs. *Else beginnt zu lachen*, kein PV). In Engerer (1999a, in Vorb.) wird gezeigt, daß die Abspaltung bestimmter Verwendungsweisen als PV, wie man sie in den genannten Grammatiken findet, nicht haltbar ist und letztlich dazu führt, die einheitliche semantische Leistung dieser Verbkategorie zu übersehen.

Dieser Beitrag soll nun u.a. auch die Position, PV seien auf infinitivische Komplemente beschränkt, hinterfragen, denn es gibt gerade fürs Deutsche gute Gründe, auch nominale Komplemente als Argumente eines PV-Prädikats im engen Sinne zu betrachten. Wenn man die Hauptfunktion von PV u.a. darin sieht, die Zeitstruktur von Sachverhaltsrepräsentationen zu 'modifizieren' (s.u.), von denen sie selbst nicht Bestandteil sind, dann ist diese Leistung in Konstruktionen mit infinitivischen Komplementen, die ja kategorial (unvollständige) Sachverhalte mit einer minimalen Zeitstruktur (Aktionsart) denotieren, unmittelbar einsichtig; es erfordert einen zweiten Blick, auch bei Nominalen eine solche intervallbezogene Modifikationsleistung zu sehen, einfach deswegen, weil Nominale kategorial nicht auf zeitlich strukturierte Dinge festgelegt sind (zu einer formalen Explikation dieser Modifikationsrelation mit wahrheitsfunktional definierten PV-Operatoren siehe Engerer (1999b, in Vorb.)). Nichtsdestoweniger zeigen PV-Infinitiv- und PV-NP-Konstruktionen Parallelen, die auf eine einheitliche Behandlung der PV-Semantik hindeuten. Dazu gehört natürlich, auch für nominale Denotate Aktionsarten in Betracht zu ziehen; ein entsprechender Ansatz wird in Engerer (1999b, in Vorb.) entwickelt.

Die Tatsache, daß PV unmittelbar auf aktionsartlichen Merkmalen ihrer verbalen Komplemente operieren, ist spätestens seit Vendler (1967) bekannt.

der die Verknüpfungsrestriktionen von *finish* und *stop* als (ein) Kriterium für die Aufteilung des englischen Verb-Wortschatzes in Aspektklassen einführte.[1] Seitdem werden PV, und darunter fast ausschließlich intervallabschlußdenotierende, als Testeinheiten für Aktionsarten benutzt, ohne daß allerdings ihre Semantik insgesamt bisher ausreichend reflektiert worden wäre. Benennungen wie '"Aspectual" Complement Verbs' (Dowty 1990: 68) oder 'Super-lexical morphemes' (Smith 1997: 48) sollten nicht darüber hinwegtäuschen, daß hier im Grunde genommen oft - und ich übertreibe jetzt bewußt - mit ad hoc-Kollokationen der Art *Hund* : *bellen* hantiert wird, deren Status in Bezug auf Aussagen zur Aktionsart von Ausdrücken alles andere als geklärt ist. Während nämlich Tests, die mit Zeitadverbialen oder dem Kriterium der Imperativbildung arbeiten, plausible situationale Korrelate in Merkmalen wie 'durativ' bzw. 'dynamisch' haben (letzteres vermittelt durch 'Agentivität', vgl. Smith (1997: 39f)), ist ein solcher Bezug bei PV undurchsichtiger: Einerseits beziehen sich PV auf 'Durativität', indem sie sie voraussetzen, andererseits fundieren sie durch die Subkategorie der intervallabschlußdenotierenden PV das dritte Merkmal, 'telisch'; im Englischen leistet dies der Kontrast *finish* vs. *stop*, wobei ersteres Verb nur telische Komplemente nimmt ('completion'), letzteres sich dagegen auch mit atelischen Komplementen verbindet ('termination', vgl. Smith (1997: 43)).

Deutet man 'completion' und 'termination' nicht als bloße Paraphrasen der idiosynkratischen Bedeutungen der englischen Verbausdrücke *finish* und *stop*, sondern als übereinzelsprachliche semantische Korrelate u.a. telischer Situationsmerkmale - wie das offenbar auch Smiths Intention ist -, sollten die entsprechenden PV auch anderer Sprachen in ähnlicher Weise sensitiv gegenüber der telisch/atelisch-Unterscheidung sein. Hierzu folgen nun sehr kursorische und vorläufige Überlegungen fürs Deutsche, die sich auf die in diesem Zusammenhang relevante Gruppe der intervallabschlußdenotierenden PV um *aufhören* und *beenden* konzentrieren. Eine modelltheoretische Darstellung von PV aller drei Intervallbereiche (linker Intervallrand (*beginnen, einsetzen* ...), Intervallmitte (*unterbrechen, fortsetzen* ...) und rechter Intervallrand (*beenden, abbrechen* ...)) ist ebenfalls Gegenstand von Engerer (1999b, in Vorb.).

Während die englischen PV fast unbeschränkt die Einbettung verbaler Komplemente erlauben, nehmen die deutschen PV, soweit sie überhaupt zweistellig sind (vgl. einstelliges *losgehen* in *Das Unwetter geht los*), überwiegend nominale Komplemente (eine Ausnahme im Intervallmittebereich, *fortfahren*). Die intervallabschlußdenotierenden PV im Besonderen zeichnen sich dadurch aus, daß sie alle, außer *aufhören*, obligatorisch nominale Komplemente nehmen, sofern sie zweistellig verwendet werden (*beenden, abbrechen, einstellen, (be)schließen*). Auch bei *aufhören*, das als einziges in dieser Gruppe infinitivische Komplemente zuläßt, scheinen nominale Komplemente nicht ausgeschlossen zu sein, besonders dann, wenn es sich um Verbnominalisierungen handelt (vgl. *Peter hörte ?die Arbeit/das Rauchen auf*). Letztendlich bleibt, will man Aktionsart weiter auf verbale Syntagmen beschränkt wissen, lediglich *aufhören*, das als einziges intervallabschlußdenotierendes PV syntaktisch für diese Kategorie zugänglich ist. Die im Englischen üblichen Tests für Aktionsarten mit intervallabschlußdenotierenden PV (*finish/stop*) sind also im Deutschen schon aus syntaktischen Gründen nicht ohne weiteres nachzuvollziehen.

Will man, um den Punkt zu demonstrieren, den durch *John stopped painting the picture* (Dowty 1990: 57) beschriebenen Sachverhalt, daß der (aspektuell unvollendet markierte accomplishment-) Teilsachverhalt 'Johns Malen eines bestimmten Bildes' von John unterbrochen wurde (und der nichts impliziert darüber, ob das Bild fertig ist oder nicht), im Deutschen durch eine streng parallele[2] Konstruktion wiedergeben, kommt man zu einem zweifelhaften ?*John hörte auf, das Bild zu malen.*[3] Auch die iterative Reinterpretation des eingebetteten Komplements als activity ('multiple-event Activity', vgl. Smith (1997: 50)) ist auf Grund des effizierten singularischen Objekts blockiert. Die Voraussetzungen für iterierte Uminterpretationen, und damit - so scheint es - für die Verbindbarkeit von *aufhören* mit 'eigentlichen' accomplishments (s.u.), sind erst gegeben bei affiziertem singularischen Objekt (*John hörte auf, seine Frau zu malen*) sowie bei pluralischen effizierten Objekten ohne Artikel und singularischen Massentermen (*John hörte auf, Bilder/Apfelmus zu malen*) (zu einer formalen Analyse von Massenterm-Prädikat-Verbindungen vgl. Krifka (1989)). *aufhören*, das sich im übrigen ohne die Notwendigkeit einer Iterierung mit activities und states verbindet (*John hörte auf, den Großvater zu quälen/ärgerlich zu sein*), erzwingt Iterierung bei accomplishments und macht sie damit zu 'sekundären' activities. Anders ausgedrückt, *aufhören* verlangt atelische bzw. konzeptuell atelisierbare Komplemente, der Satz von Dowty kann also unter den strengen Identitätsbedingungen, s. Anmerkung 2, eigentlich nicht übersetzt werden.

Welche Ressourcen hat das Deutsche, unter weniger strengen Bedingungen, das Aufhören eines accomplishment-Sachverhalts auszudrücken? Eine naheliegende Möglichkeit besteht darin, das infinitivregierte Akkusativobjekt in eine *an*-oder gegebenenfalls *in*-Präpositionalphrase umzuformen (*John hörte auf, am Bild zu malen/im Buch zu lesen*, vgl. zu dieser Konstruktion Filip (1989) sowie Engelberg (1994)). Folgt man Smith (1997), entsteht ein activity. Man könnte allerdings auch den Standpunkt vertreten, es handle sich weiterhin um (telische) accomplishments, da die Telizität der beschriebenen Ereignistypen (in einem noch zu erläuternden Sinn, s.u.) auch in der *an*-Konstruktion rekonstruierbar bleibt. Diese Meinung vertritt Filip (1989), die dann konsequenterweise die präpositionalen Konstruktionen des Typs *an* als Markierung des imperfektiven Aspekts deutet. Ich neige zur letzteren Ansicht. Verknüpft man nämlich einen Adverbialsatz wie ... *bis X fertig war*, der die Telizität (eben nicht die Perfektivität!) seiner satzförmigen Bezugskonstituente verlangt, mit einer *an*-Konstruktion, zeigt sich, daß die Akzeptabilität der Gesamtkonstruktion mit der, im Sinne Filips, 'zugrundeliegenden' Telizität im Bezugssatz positiv korreliert. Der Adverbialsatz verbindet sich deutlich besser mit telischen als mit atelischen Vorgängern, ungeachtet ihrer aspektuellen Imperfektivität. Vgl. *John malte am Bild, bis er/es fertig war* vs. **Petra fühlte am Stoff, bis *sie/*er fertig war* (näheres in Nicolay/Engerer (1999, im Druck); dort werden auch die jeweils präferierten Referenzindizes für referentiell zweideutige Pronomen im Adverbialsatz diskutiert).

Eine weitere Möglichkeit im Deutschen, 'accomplishments aufzuhören', besteht darin, das syntaktische Objekt des Infinitivverbs oder das gesamte infinitivische Komplement zu tilgen, also elliptisch zu konstruieren. Im ersten Fall entsteht ein syntaktisch eingebettetes activity, im zweiten Fall ein achievement (*aufhören*) ohne syntaktisches Komplement (A sieht B beim Malen eines Bildes zu. A zu B: 'Wann hörst du endlich auf (zu malen)?'). Solche Verwendungen sind äußerst kontextabhängig und zeigen nur noch einmal, wie rigide die aktionsartliche Grammatik von *aufhören* ist. Konzeptuell spricht nämlich nichts dagegen, den Begriff des Aufhörens auf telisch beschriebene Sachverhalte anzuwenden. Die einzige konzeptuelle Beschränkung von *aufhören* liegt darin, daß der vom Komplement denotierte Bezugssachverhalt nicht als abgeschlossen markiert sein darf (alles andere ist einzelsprachliche Grammatik). Anders gesagt, aufgehört werden kann etwas nur, wenn es von vornherein keinen angelegten Endpunkt aufweist, oder, falls ein Endpunkt angelegt ist, es diesen Endpunkt nicht erreicht hat. Der Begriff des Aufhörens, nicht die Grammatik des deutschen *aufhören*, stellt also Bedingungen an das tatsächliche Erreichen eines Endpunkts, die konzeptuelle Beschränkung ist damit aspektueller Art. Aktionsart hängt mit Aspekt lediglich insofern zusammen, als nur telische Prädikate perfektiv sein können, da sie - und nur sie - den Endpunkt lexikalisch quasi 'anbieten' (Filip (1989); vgl. dagegen Smith (1997), wo auch atelische Prädikate, die u.a. mit Zeitadverbialen des Typs *von 13 Uhr bis 14 Uhr* verbunden werden, dem perfektiven Aspekt zugerechnet werden). In Aspektsprachen wie dem Russischen spiegelt sich diese begriffliche Beschränkung in der Grammatik wider, indem die Gruppe der intervallabschlußdenotierenden PV obligatorisch imperfektiv markierte Verben einbettet, ungeachtet ihrer aktionsartlichen Eigenschaften. Im Deutschen dagegen, das über keine systematische Aspektmarkierung verfügt, sind Telizität und Perfektivität aneinander gekoppelt, wenn die Ausdrücke als infinitivische *aufhören*-Komplemente erscheinen.[4]

Egg (1994: 51) bemerkt diese Probleme in seinem Buch, wenn er sagt, *aufhören* sei nur auf unbegrenzte, das sind bei ihm u.a. (atelische) activities oder states, anwendbar. Er formuliert deswegen den, wie er es ausdrückt, 'ursprünglichen' Test mit *stop* aus Dowty in einen Test mit dem deutschen PV *unterbrechen* um, das zwar die (im Gegensatz zu *aufhören*) gewünschten semantischen Eigenschaften von *stop* habe, aber obligatorisch mit nominalen Komplementen konstruiert wird. Auch unter diesen syntaktischen Voraussetzungen kommen nach Egg aktionsartliche Unterschiede zum Vorschein, z.B. solche zwischen accomplishments und achievements: Während accomplishments als *unterbrechen*-Komplemente keine iterative Reinterpretation verlangen (*Ariadne unterbrach das Schreiben ihrer Antrittsvorlesung*), ist dies bei achievements die bevorzugte Interpretation

(*Ariadne unterbrach das Gewinnen des Rennens*). PV modifizieren also keineswegs nur verbal denotierbare Sachverhaltsrepräsentationen, sondern auch - und in sehr ähnlicher Weise - nominale (im obigen Fall ein einfaches Verbalsubstantiv mit Genitivattribut, das den Endpunkt des 'nominalen accomplishment' denotiert).

Zu fragen bleibt, wie man von der aktionsartlichen Klassifizierung eines verbalen Syntagmas *die Antrittsvorlesung schreiben* zur identischen Klassifizierung des nominalisierten Gegenstücks *Schreiben der Antrittsvorlesung* kommt. Dieses Problem stellt sich verschärft bei *beenden*, dem Pendant des englischen 'completion'-Verbs *finish*. Letzteres schließt, wie bereits gesagt, schon syntaktisch die Einbettung infinitivischer Komplemente aus, ungeachtet dessen, ob das Infinitiv-Objekt realisiert ist oder nicht (**Peter beendete (das Bild) zu malen*). Während also 'termination' eines accomplishment durch *aufhören* sozusagen zur Ellipse des eingebetteten Objektausdrucks führt (*Peter hörte auf, (?das Bild) zu malen*, vgl. oben), wird ein 'completion'-Sachverhalt im Deutschen, wenn man so will, verbelliptisch konstruiert, indem (hier: valenzobligatorisch) nur das effizierte Objekt syntaktisch zum Vorschein kommt: *Peter beendete das Bild (*zu malen)*. Der verbale Typ des eingebetteten Prädikats Präd in Präd (Peter', (das Bild)') ist damit in der Konstruktion mit *beenden* syntaktisch unspezifiziert, bleibt allerdings latent vorhanden, da seine Rekonstruktion für die Bedeutungszuweisung des gesamten *beenden*-Satzes notwendig ist (dasselbe gilt für andere mögliche 'completion'-Synonyme wie *vollenden* oder *abschließen*).

Die Struktur eines *beenden*-Satzes mit zweistelligem Präd kann in einer ersten Näherung angegeben werden als

beenden' (X, $\text{Präd}_{AA}$ (X/Y, $\text{Thema}_{KL}$))

X, das syntaktische *beenden*-Subjekt, kann erstes Argument beider Prädikate sein (*$X_i$ beendete seinen$_i$ Lauf*, X-X, Koreferenz), oder referentiell auf die erste Stelle des übergeordneten beenden' beschränkt bleiben (*X beendete den Lauf seines Freundes*, X-Y, Y 'externer' Referent von Präd). KL ist eine hybride Klassifizierung der Thema-Denotate bzw. Thema-Ausdrücke (syntaktische *beenden*-Objekte) in a) zeitlich nicht-strukturierte Gegenstände wie Bilder, Steine und Franz, b) zeitlich strukturierte Gegenstände wie Katastrophen oder Tage, deren Ausdrücke im Deutschen keinen morphologischen Bezug zu Verben unterhalten (*Katastrophe*, *Tag*) und c) die morphologische Klasse von Verbalsubstantiven im weitesten Sinn wie *Lauf*, *Versammlung*, oder *Unterricht*. KL bestimmt u.a. die Aktionsart AA von Präd, die thematische Markierung des ersten Arguments X/Y und des Themaarguments sowie die Koreferenzbeziehungen X-X/Y. Fällt das Thema beispielsweise unter die KL-Subklasse a), wie in *X beendete das Buch*, ist Präd als zweistelliges accomplishment mit einem Agens als erstem und einem effizierten oder affizierten zweiten Argument (konsumierte Argumente sind ausgeschlossen) zu rekonstruieren, wobei Koreferenz X-X erzwungen ist: beenden' (X, Präd$_{ACC, 2ST}$ (X$_{AG}$, Buch'$_{EFF, AFF}$))

Auf ganz parallele Weise steuern auch die anderen KL-Subkategorien die Festlegung der für die Bedeutungsermittlung notwendigen Merkmale. Präd selbst, genauer gesagt: der lexematisch-idiosynkratische Bedeutungsbestandteil, denotiert eine Menge von Prädikaten, die in Abhängigkeit von KL und dem Themadenotat definiert wird. Die Regel beispielsweise für KL = c) mit Diskussion als Themadenotat ist relativ einfach (*X beendete die Diskussion*): In diesem Fall enthält die Menge Präd ein Element, ‚diskutieren', das in Abhängigkeit von KL (morphologische Ableitungsbeziehung) Präd zugewiesen wird. Es ergibt sich:

beenden' (X, diskutieren'$_{ACT}$ (X/Y))

Zu sehen ist hier auch, daß weder Stelligkeit noch Aktionsart (hier: activity) des eingebetteten Prädikats beschränkt sind, auch 'externe' Referenz Y ist zugelassen.

Ist KL = a), liegt für Präd im Normalfall eine mehrelementige Menge vor. Sie besteht aus Prädikaten, die die KL-Bedingung a) erfüllen (zweistelliges accomplishment, X ist Agens usw., s.o.) und Relationen des Typs a) zwischen X und dem Themadenotat dergestalt aufbauen, daß sie aus dem Kontext oder dem Weltwissen über konventionelle Relationen bezüglich bestimmter Dinge rekonstruierbar sind. Die letzte Formulierung ist (noch) vage, aber es ist klar, daß die Mengen von zu rekonstruierenden Prädikaten jeweils beschränkt ist durch das Themadenotat, vgl. *X beendete das Bild*, wo Präd Elemente wie malen', zeichnen' o.ä. enthält, aber eine Interpretation durch verstecken', zerstören' usw. ausgeschlossen ist.

Der Rest der Überlegungen konzentrieren sich auf einige Analyseprobleme im Rahmen von KL-a)-Objekten, also Nominalen wie *Buch* oder *Bild*. Daß konsumierte Themaargumente ausgeschlossen sind, wurde bereits erwähnt. *X beendete das Glas/das Schnitzel* kann nicht als in beenden' eingebettetes (aus)trinken' (X, (das Glas)') bzw. (auf)essen' (X, (das Schnitzel)') interpretiert werden. Eine weitere Eigenschaft solcher Konstruktionen besteht darin, daß eine imperfektive Lesart von Präd-accomplishments blockiert zu sein scheint: *X beendete das Buch*$_{i,\ *EFF/?AFF}$ *und *schrieb*$_{EFF}$*/?las*$_{AFF}$ (*es*$_{i,\ EFF/AFF}$) *am nächsten Tag weiter*

Während die imperfektive Interpretation des ersten Teilsatzes, etwa beenden' (X, schreiben-an'/lesen-in' (X, (das Buch)'), bei affizierendem *lesen* vielleicht möglich ist, erscheint sie bei effiziertem Objekt ausgeschlossen. Damit hängt eine weitere Beschränkung zusammen, die auf

Wahrheitsintervallen bezüglich identischer Referenten operiert. Ein eingebettetes Präd mit koreferentem X an der Agensstelle umfaßt bei Effiziertheit (schreiben', malen' usw.) nicht notwendigerweise das gesamte accomplishment als Wahrheitsintervall, sondern nur Schlußphase inklusive Endpunkt. Mit anderen Worten, die koreferente Agensbesetzung mit X (d.h., X ist erstes Argument sowohl von beenden' als auch von Präd) ist obligatorisch für ein finales Teilintervall des accomplishment, Initialphase sowie folgende interne Stadien können auch andere Referenten Y involvieren.[5] Ein Satz wie *X beendete das Buch* mit der effizierten Interpretation von Präd als schreiben' folgert nicht, daß X in allen Phasen des accomplishment als Agens fungiert, insbesondere nicht für die Initialphase. Folgender Satz, obwohl stilistisch kein Meisterwerk, ist kein Widerspruch:

$X_i$ *beendete das Buch*$_j$, EFF, *obwohl er*$_i$ *es*$_j$ *nicht begonnen hatte (zu schreiben)*

Der erste Teilsatz ist paraphrasierbar etwa als 'X schrieb das Buch fertig'. Affizierte Präd dagegen schließen die Initialphase als Bestandteil des Wahrheitsintervalls mit koreferentem X ein ('initiales Teilintervall', vgl. Lohnstein (1996: 227)). Etwas salopp ausgedrückt, ein X kann ein Buchlesenereignis, in dem es selbst als Agens fungiert, nicht 'wahrheitsgemäß' beenden, ohne auch als Agens in der initialen Phase des Ereignisses involviert gewesen zu sein. Vgl. wiederum den Satz *X beendete das Buch*, nun allerdings mit einem affizierten Präd lesen'. Die parallele Sequenz zu oben führt jetzt zu einem Widerspruch ($X_i$ *beendete das Buch*$_j$, AFF, *obwohl er*$_i$ *es*$_j$ *nicht begonnen hatte (zu lesen)*).[6]

Sätze wie *Ein Unglück beendete das Buch* oder *Ein Strich beendete das Bild* scheinen der oben aufgestellten Agensbedingung für die Subjektdenotate zu widersprechen. In einer Lesart, nämlich 'ein Unglück (d.h. die Schilderung eines Unglücks) schloß das Buch ab', findet eine Umkonzeptualisierung von

BUCH[7] von einem zeitlosen, konkreten Gegenstand zu einem Bündel von geschilderten, zeitlich strukturierten Handlungen statt, etwa analog zu 'Ein Unglück beendete die Schule', wo SCHULE zu einem sich regelmäßig wiederholenden institutionellen Handlungsmuster umkonzeptualisiert wird (Bierwisch 1983). Solcherart umkonzeptualisierte *beenden*-Objekte führen zu einem Übergang in KL-b), die als Kernklasse zeitstrukturdenotierende Ausdrücke wie *Feier* oder *Jahr* enthält. (Parallel ließe sich bei STRICH und BILD argumentieren, nur daß hier die räumliche Organisation des Bildes auf eine zeitliche übertragen sowie ein zeitindexiertes Strichereignis konzeptualisiert wird, das innerhalb der Zeitstruktur des Bildes am Skalenendpunkt lokalisiert ist). Der Hauptpunkt ist aber, daß solche Sätze aufgrund der Agensunfähigkeit ihrer Subjektdenotate ein ungenanntes Agens involvieren können und damit zu einer zweiten, allerdings jeweils unterschiedlich akzeptablen Lesart führen. Vgl. die Paraphrasen 'X beendete das Buch damit/dadurch, daß er ein Unglück erlitt' vs. 'X beendete das Bild damit/dadurch, daß er einen Strich zog'. Die Verhältnisse sind hier nicht ganz klar. Es scheint aber, daß neben Präd und beenden' eine dritte Prädikation involviert ist, hier Unglück-erleiden' (X/Y) bzw. Strich-ziehen' (X/Y), die ein finales Teilintervall des beenden'-Sachverhalts charakterisiert.[8] Während es nun das Denotat von *Strich* erlaubt, ein Ereignis zu konzeptualisieren, in dem X in der Agensrelation auftritt, ist dies in Bezug auf ein Unglücksereignis mühsamer, da dem Mitspieler im ersten Zug die Patiensrolle zukommt. Diese fehlende Agentivität könnte dafür verantwortlich sein, daß die oben paraphrasierte zweite Lesart für den Unglück-Satz fragwürdig ist. Aber hier ist noch eine genauere Analyse notwendig.

Eine spezielle Behandlung erfordern Fälle wie ?*X beendete die Beule*, wo die zu rekonstruierenden Prädikate achievements wie sich-zuziehen' oder sich-schlagen' sind[9] und somit gegen die accomplishment-Bedingung verstoßen.

Dazu kommt, daß unter referenzidentischer X-X-Interpretation X nicht als Agens eines Ereignisses, bei dem es sich selbst die Beule zufügt, konzeptualisiert werden kann (X wäre gleichzeitig Patiens).

Diese letzten, sehr vorläufigen Bemerkungen sollten zeigen, wie man sich eine Erweiterung von zeitlicher Semantik auf Nominale in bezug auf PV in etwa vorstellen könnte und welche weiteren Forschungsfragen damit einhergehen. Daß damit mehr Probleme aufgeworfen sind als gelöst, ist unbestritten. Deutlich sollte aber geworden sein, daß Sätze des Typs *beenden* nicht verstehbar sind, wenn ihre nominalen Komplemente nicht als satzartige, d.h. zeitlich strukturierte Einheiten mit Individuen als Mitspielern rekonstruiert (hier: repräsentiert) werden, und daß jene wiederum aktionsartlichen Beschränkungen in Bezug auf PV unterliegen. So wird es vielleicht möglich, eine nicht in Varianten aufgespaltene Semantik der PV zu erarbeiten, die sich einheitlich auf sententiale sowie nominale Komplemente bezieht.

*The Department of German, University of Aarhus*
*Ndr. Ringgade, bygn. 326*
*DK-8000 Århus C, Denmark*

## Anmerkungen

[1] Tatsächlich klassifizierte Vendler Verbalphrasen nach ihrer Aktionsart. Die Bezeichnung 'Aspekt' ist u.a. Dowtys 'aspectual classes' zu verdanken, der sich allerdings der terminologischen Probleme bewußt ist, vgl. Dowty (1990: 52); zur Diskussion speziell der Aspekt-Aktionsartproblematik siehe Nicolay/Engerer (1999, im Druck).

[2] 'Parallel' heißt: a) Ausdrücke für dieselben Individuen (John, das Bild) werden eingeführt und die Repräsentationen denotieren b) identische Relationen (R1 = malen' (John', (das Bild)')), R2 = unterbrechen' (John', R1)) sowie c) identische Aktionsarten (accomplishment (R1), achievement (R2)).

[3] *John unterbrach das Malen des Bildes* mit Nominalisierung genügt den in Anmerkung 2 genannten Bedingungen nicht, weil die Korrespondenz der Verbnominalisierung mit Genitivattribut zu R1 infolge der referentiellen Zweideutigkeit der Besetzung des ersten Arguments in Bezug aug R2 nicht gegeben ist; zudem ist die Aktionsartenzuweisung an Nominale nicht geklärt, vgl. Bedingung c).

[4] In anderen syntaktischen Konfigurationen, beispielsweise in Satzgefügen, ist diese Kopplung selbstverständlich nicht gegeben, vgl. *als er aus dem Büro nach Hause zurückkehrte, regnete es*, wo die Telizität des *als*-Satzes die imperfektive Interpretation keineswegs blockiert.

[5] Zur Definition verschiedener Teilintervalle s. Lohnstein (1996: 224ff) sowie ausführlicher zur Intervallsemantik Dowty (1990: 133ff).

[6] Der rechte Endpunkt eines affizierten accomplishment-Prädikats Präd ist offenbar nicht obligatorisch im Wahrheitsintervall von beenden' enthalten. Er kann, unter geeigneten Kontextbedingungen, sozusagen neutralisiert werden, z.B. in *X beendete das Buch nach ein paar Seiten/nach der Hälfte*. Die effizierte Interpretation ist, wie schon erläutert, ausgeschlossen.

[7] Kapitälchen signalisieren hier Konzepte.

[8] Die kausal-instrumentale Relation ist hier vernachlässigt. Eine Ausdehnung des charakterisierenden Sachverhalts auf eventuelle Folgephasen von Präd ist in dieser Konstruktion übrigens nicht möglich. Erlaubt ist dies bei KL-c)-Objekten wie in *Ein kleiner Umtrunk beendete die Diskussion*.

[9] *sich eine Beule zuziehen/schlagen* sind punktuelle achievements, was man u.a. daran sieht, daß sie sich auf einen unmittelbaren Nachzustand beziehen, wenn sie mit bestimmten Zeitadverbialen verbunden werden, vgl. *Franz zog sich 2 Stunden eine Beule zu*.

## Literatur

Bierwisch, Manfred. 1983. Semantische und konzeptuelle Repräsentation lexikalischer Einheiten. In: Untersuchungen zur Semantik, ed. by Rudolf Ruzicka and Wolfgang Motsch, 61-99. Berlin: Akademie-Verlag.

Dowty, David R. 1990. Word meaning and Montague Grammar. Dordrecht, Boston, London: Reidel.

Egg, Markus. 1994. Aktionsart und Kompositionalität. Berlin: Akademie Verlag.

Engelberg, Stefan. 1994. Valency and Aspectuality. Syntactic and Semantic Motivation for the Notion of 'Change of State'. In: Sprache - Sprechen - Handeln. Akten des 28. Linguistischen Kolloquiums, Graz 1993, Vol. 1, ed. by Dieter W Halwachs and Irmgard Stütz, 53-59. Tübingen: Max Niemeyer Verlag.

Engerer, Volkmar. 1999a, in Vorb. Infinitivische Komplemente von Phasenverben in deutschen Grammatiken. To appear in: Akten des 33. Linguistischen Kolloquiums, Thessaloniki 1998, ed. by Dorfmüller-Karpusa. Frankfurt am Main: Lang.

Engerer, Volkmar. 1999b, in Vorb. Towards a model theoretic explication of phase operators in a Montague Grammar framework.

Filip, Hana. 1989. Aspectual Properties of the AN-Construction in German. In: Tempus - Aspekt - Modus. Die lexikalischen und grammatischen Formen in den germanischen Sprachen, ed. by Werner Abraham and Theo Janssen, 259-292. Tübingen: Max Niemeyer Verlag.

Helbig, Gerhard, and Joachim Buscha. 1984. Deutsche Grammatik. Leipzig: VEB Verlag Enzyklopädie.

Krifka, Manfred. 1989. Nominalreferenz, Zeitkonstitution, Aspekt, Aktionsart: Eine semantische Erklärung ihrer Interaktion. In: Tempus - Aspekt -

Modus. Die lexikalischen und grammatischen Formen in den germanischen Sprachen, ed. by Werner Abraham and Theo Janssen, 227-258. Tübingen: Max Niemeyer Verlag.

Lohnstein, Horst. 1996. Formale Semantik und natürliche Sprache. Opladen: Westdeutscher Verlag.

Mulisch, Herbert et al. 1985. Die russische Sprache der Gegenwart. Vol. 2: Morphologie. Leipzig: VEB Verlag Enzyklopädie.

Nicolay, Nathalie, and Volkmar Engerer. 1999, im Druck. Telizität, Aspekt und Zustandswechsel. To appear in: Akten des 32. Linguistischen Kolloquiums, Kassel 1997, ed. by Hans-Otto Spillner and Ingo Warnke. Frankfurt am Main: Lang.

Smith, Carlota. 1997. The Parameter of Aspect. Dordrecht, Boston, London: Kluwer Academic Press.

Vendler, Zeno. 1967. Verbs and Times. In: Linguistics in Philosophy. Chap. 4, 97-121. Ithaca, New York: Cornell University Press.

Zifonun, Gisela et al. 1997. Grammatik der deutschen Sprache. 3 Vols. Berlin, New York: Walther de Gruyter.

# Some reflections on syntactic structure and the acquisition process.

*Peter Harder*

### Abstract

The paper presents a case for the importance of syntactic structure while maintaining that structure is not autonomous, but function-based. The claim combines elements from the familiar opposed standpoints, where structure, binary contrast, rationalism and nativism is one and function, gradualism, empiricism and nurture is the other. Syntactic structure is argued to be central, but its genesis as well as its ultimate role in language depends on functionalist pressures operating both within the cognitive developmental process and from the social context of learning, operating in a Darwinist manner. Some observations are offered on the nature of the cohabitation between probabilistic and lexical features of language on the one hand and structural and syntactic facts on the other.

## Introduction

The following paper argues for a recognition of the special properties of syntax within an overall functionalist approach, in which language as a social fact is basic. This alliance between structure and social selection is motivated by an assumption that the existence of a learning target with particular structural properties, a goal of learning felt as compelling by the language-acquiring child, is due to social forces. The idea is that the probabilistic, item-based, bottom-up elements in language learning are channelled by a top-down mechanism which makes structural features emerge from them without being reducible to them. What is necessary to go from the probabilistic hit-or-miss stage to full membership of the speech community is thus neither the maturation of pure innate structure nor learning any definitive instalment of language patterns, but rather a kind of fossilization that attunes the production and processing patterns to just that set of structural regularities that are authorized by the speech community.

That does not imply that the probabilistic elements are put out of business - they go on working and are powerful enough to cause speakers to produce utterances that do not conform to structural patterns. However, unless speakers were aware of the structural relations there would be no sense in which such utterances were violations. In this picture, speakers' personal 'competences' are essentially approximations to the structural norm of the speech community — just as internalized mental models of *e.g.* city maps are approximations to actual facts about cities.

## Function-based syntax: an illustration

The approach I am arguing for aims to capture some of the same facts that are often taken to support the existence of autonomous syntactic structure — but tries to found them on function-driven processes. The first aspect of the argument depends on a scale of cognitive competence, in terms of which syntactic hierarchy occupies a distinct level of achievement. The extra level occupied by syntax can be viewed both top-down and bottom-up, the latter being the standard approach; I try to argue that it is just as revealing to see syntax as the result of a differentiation of utterance meaning into subparts as it is to see it in terms of the emergence of combinations of words.

As an example one can take the sequence of acquisition of negation, which is often put forward as mirroring an innate developmental logic, cf. e.g. Felix (1987: 98). If we include the one-word stage, the commonly assumed stages are (1) NEG! (2) NEG + utterance; (3) Subject + NEG + rest-of-utterance; (4) Subject + AUX/NEG + rest-of-utterance. However, if syntax is seen as based on functional relations between meaningful elements, the sequence described above can be plausibly accounted for in terms of a different scenario. The argument is based on an analogy between language and tool use described by Greenfield (1991: 533). She outlines three stages of competence, which for my purposes I have illustrated as follows:

(1) single act/holophrase (no internal structure). Tool use example: the subject grabs an object (e.g. a spoon). Linguistic example: uttering a holophrase, such as *more!*

(2) pairing (two acts, one after the other). Tool use example: grabbing a spoon, then putting it into the mouth OR grabbing a spoon and putting it into the food. Linguistic example: *more cookie,* OR *cookie more.*

(3) hierarchy (imposing a two-stage structure on the sequence of acts - which simultaneously makes the order meaningful). Tool use example: grabbing a spoon, then putting it into the food - and THEN putting the spoon into the mouth. Linguistic example: *want (more cookie).*

I would like to use the scenario to highlight an important feature of syntactic structure, namely that it requires freedom from stimulus control (as emphasized in Chomsky's original attack on behaviourist assumptions about language). In the case of spoon-handling, the third, hierarchical stage is only achieved if the subject can organize the sequence, which entails that the act of putting the spoon into the mouth is postponed until the preconditions are met (i.e. when there is food on the spoon). It is only this organization which makes the sequence as a whole meaningful in spite of the fact that the first two acts do not make sense in themselves (in the sense that they could not be attributed to direct stimulus control). Similarly, syntactic planning presupposes that the subject can use each word independently of stimulus control in order to create an utterance which as a whole makes the right kind of sense. The ability to discipline the 'firing' of individual impulses and impose an ordered pattern of behaviour upon them is crucial.

Applied to the case of negation, the same logic explains why the first step is a holophrase (*No!)*, without syntactic structure, which applies directly to whatever situational stimulus is at hand. The next step is the ability to pair negation with an indication of what is to be negated: *nappy no* or *no nappy*. The third step is to create a hierarchy: *mummy no go*. The analysis applies to the content, or 'function' aspect of the words, regardless of exactly how the hierarchy is signalled; but regularities on the side of expression, or 'form' have their own role to play (cf. Lund 1996). The fourth step cited above is essentially a matter of the English expression pattern rather than a special hierarchical step.

The route that is indicated in this example is the traditional bottom-up path from single elements to structured combinations of elements. However, I would like to stress the top-down element in the picture. The point can be seen if one takes into consideration that all three example utterances essentially mean the same thing in the right context. From that point of view the path is therefore not from isolated elements to coherent combinations — it is rather from a primitive and more context-bound way of signalling your intentions to more complex and less context-dependent ways. From that point of view, the essential development is the 'downward' differentiation of a total message into independently codable subparts (cf. also Harder 1996: 219) — rather than the 'upward' combination of single items into a complex message. In terms of language learning, this reflects the path from the baby's use of one imperious all-purpose holophrase *(AA!)* to get the parents to do the required thing, to more differentiated means of expression.

In terms of this picture, syntactic learning is a pincer operation: it moves 'downwards' by differentiating holophrastically signalled utterance meanings into subparts, thus gradually decreasing the dependence on contextual stimuli of utterance understanding, while at the same time moving 'upwards' by enabling the combination of individual items into more complex constituents.

## The ontology of structure — and of learning

The progress of this pincer movement is subject to very great variation both between individuals and in any other dimension. From a cognitive point of view it is natural to see structure as something that gradually emerges in the learner - whether caused by innate predisposition (which is perfectly compatible with the scenario above, and orthogonal to the issue of syntactic autonomy) or features of the communicative environment. But I would like to throw this way of looking at it into relief by seeing structure initially as something existing outside the learner: to the extent language learning is a form of adaptation to the environment, structure is a feature of the learning target, and not exclusively of the mind-internal process.

The language-learning child is engaged in a process that can plausibly be described as 'cracking the code' — looking for whatever regularities make coherent sense of the complex utterances that the child hears around her, first in order to understand, later in order to use them. In this light, emerging structural features in the child's language (receptive as well as productive) appear as downloaded copies of items from the language net that the child is plugging into - copies distributed without the benefit of a central server, as David Tuggy once put it. Structure exists as something to be mastered, like a mountain that is there to be climbed. The child, thus construed, does not create the structure, any more than the climber creates the mountain: what both do is to make their own way to a target that is there, in the environment. In doing so, children produce mental structures of their own (without which language as a social construct could not be kept going); but these structures are not the sole ontological locus of linguistic structure.

Obviously, the target is less well-defined in the case of the language learner than in the case of the climber. The child, as often pointed out, has to generate her own language(-copy) from input that does not specify the relevant properties exactly. Hence, there is room for individual differences even at the stage of target competence, with all that that entails. Looking at actually occurring language data therefore never gives you any precisely specified grammatical structure: language as an empirical range of phenomena is essentially probabilistic.

This may throw some light on the familiar discussion about whether structure is the linguist's hypostatization of something that is essentially gradual, approximative patterns of events. The point of view according to which structure is initially outside the child creates room for an either-or element over and above the approximative elements that meet the eye when one looks at the learning process. The either-or element is in the element of membership of the speech community. Initially, the child is left out of what is going on linguistically; at the target point, the child is inside — an advance from minus to plus.

This coexistence between clines and digital choices is more or less a design feature of complex systems, a feature of the relation between higher-order constructs and the foundations they build on. Extinction vs. survival of species is in some sense a yes-no issue (will there or will there not be Sumatran rhinos a hundred years from now?) — but if we describe the process of extinction empirically in real time, it is always a matter of more or less specimens, breeding couples, offspring, etc; zero is just another point on the scale of numbers. To take a cognitive example, the existence of categories (over and above instances) presupposes the existence of an either-or element in addition to clines: unless the cline of individuals is describable as stretching between ending points defined in terms of membership or not, there would be no point in categorization. In ascending from the level of instances to the level of categories, therefore, some information is lost - and that indeed is the whole point of the exercise: categorization reduces complexity to more manageable proportions. Keeping the whole extent of gradual variation at the level of categories would leave us exactly where we were before we had categories (cf. also the introduction to Givón 1995).

This duality is twofold in the case of social systems supported by individual minds. Once a social system such as a speech community or a business company is in existence, the individual is faced with a choice that can to some extent be described as either being a member of the system or not. To this social issue corresponds the cognitive issue of whether the cognitive system of the individual is capable or not of supporting the type of practise that constitutes membership of the social system. The fact that there is an either-or element, however, is fully compatible with a degree of fuzziness so that the membership as well as the cognitive competence typically lend themselves to description in terms of a cline from undisputed YES to undisputed NO. The point is that unless the plus/minus choice was there, you could not recognize the borderline as being fuzzy, because the concept of a borderline would not make sense.

Neural nets as models of cognitive capacity are good illustrations of the co-existence of gradual learning processes and (more or less) precise targets. They are trained to approximate input-output relations, and the inner anatomy of the way they manage to do it is not the most interesting thing about them; there are always many ways of doing it. This peaceful co-existence even extends to capabilities which are naturally described in terms of rule-following. As pointed out in Elman et al (1996:102), neural nets are function approximators, and "...functions are nothing if not rules".

## Syntax and the lexicon

The status of syntax vis-a-vis other aspects of knowledge of language is one of the most heated issues in contemporary linguistic theory. One implication of the top-down approach to syntax is that you cannot understand syntactic capability fully by looking at lexical items: syntax presupposes the lexicon but goes beyond it in one crucial respect. The discussion about syntax is rendered difficult by a certain lack of mutual comprehension between the parties: adherents of autonomous syntax tend to use all arguments for claiming that syntactic categories are not totally predictable from semantics as evidence that syntax is totally autonomous; anti-autonomists tend to use arguments for claiming that syntactic regularities are semantically motivated to claim that there is total continuity between lexical and syntactic properties. I would like to illustrate my view by a comparison with an attractive and widespread way of looking at syntax, that of construction grammar (cf. e.g. Goldberg 1995).

In this approach, the integratedness of syntax with the rest of language is expressed by presenting all knowledge of language in a single monostratal format, whose central component has essentially the form of a list. In this 'constructicon', each construction (the term 'construction' including minimal signs as a special case) is described with all its features, semantic, syntactic and phonological; systematic features are essentially covered by inheritance between constructions at different levels of generality.

In order to understand what is involved in this issue it is necessary to look at two ways of understanding lexical and syntactic properties. A totally non-syntactic view of linguistic competence would be one in which 'item recall' was the only skill involved. All learning would be like vocabulary learning: a gradual induction-based generalization of the suitedness of an item to indicate a certain meaning. You would use abstract syntactic constructions like the subject-verb construction the same way as you would use a lexical item like *serendipity* (which is also a construction). I assume that no-one believes this is a really good way to describe human languages.

A slightly more syntactic theory would be one in which syntax exists as an admittedly special range of properties of lexical items. Those special properties would require a special type of knowledge of the speaker — involving a command of operations enabling the speaker to link up lexical items properly. Syntax has special status, but still within the shared format of lexicon-based knowledge. I think that such a description would be able to cover all syntactic regularities. But I still think there would be something missing: the same format is used about truly lexical items, about semi-idiomatic constructions such as the *let alone* construction (cf.Fillmore et al. 1988) and about constructions without lexical content such as the subject-verb construction, implying that one size fits all.

This leaves out one half of the identity of syntax, namely the fact that syntax constitutes a level of description above 'real', prototypical lexical items. As is apparent from the top-down perspective, syntax is not only a set of elements which you choose just as you choose any other element: syntax also covers the way you are forced by the code to *organize* those elements you choose. To grasp simultaneously the contribution of items and the contribution of the way you organize them is an inherently more complex skill than just controlling the items.

To say that syntactic skill stands on the shoulders of lexical skill is to say two things simultaneously: that it is non-autonomous and that it has a distinct role of its own. Syntax regulates how to combine lexical items into full utterance meanings so that a combinatorial explosion is avoided - which would occur if all possible combinations were equally permissible. By creating a severely limited number of syntactic categories, syntax forces the richness of potential messages into a manageable format: a reduction of complexity by eliminating some of the options. There is some analogy with the building of paved roads across the landscape: without them, the landscape is more varied and one is less predetermined with respect to the route one takes - but it is so much easier to get to your destination once they are there.

## Activation and suppression

If acquisition of structure consists in establishing a norm for combination options, it is not to be expected that syntax arises in the form of innate rules breaking through the empirical mess like hard teeth through the soft gums. Rather, what happens is that a privileged set of combination operations gradually becomes more and more dominant, while co-existing at all points along the line with some of the messiness from which it is becoming distinct. The same Darwinist selection pattern that has been shown to exist in the immune system of the individual by Edelman (1992) can drive the process: combinations are produced more or less at random and then the right ones are favoured, gradually driving the wrong ones (almost) out of business. In language learning, there is a partial parallel with the phonetic part of the acquisition process: in the babbling period, children can make distinctions that they lose once it has become clear what distinctions are relevant for the language they are learning. 'Syntactic babbling' involves more or less free combinations of existing lexical items, corresponding to the 'pairing' phase in the development described above - without any hierarchically determined restrictions. The special role of syntax is the hierarchical component described above, with its privilege of conveying information bits in other ways than simply adding them on top of one another, leaving the rest to guesswork; selection pressure favours not only certain expression patterns, but also the ability to impose a reliable order upon the meaning, making it clear who did what to whom.

The parallel is only partial, however, since babbling is assumed to start out as a pre-linguistic process of trying out the vocal channel, only gradually becoming shaped by linguistic input. It is less plausible that word combinations start out as something totally untainted by structure; when 'mean length of utterance' starts to rise, part of the motivation is presumably some idea about the point of combining words, however 'proto-linguistic' it may be (cf. Bickerton 1990). What could arguably be the same is the mechanism of gradual approximation to a target.

One of the gradual factors that favour the restriction of pairing options is the increasing need to rationalize as the accumulated body of stored utterances, fragments and items becomes so large that it is an advantage to draw on it in a systematic fashion, cf. Locke (1996). It is often underlined that the human capacity to retrieve stored information is so powerful that there is no natural point at which processes of generalization and rule application 'must take over'. But no matter how great the capacity, the processing effort involved in retrieval, use and comprehension of fragments is bound to increase with the size of the inventory, and consequently the additional effort involved in systematization (whatever it may be) will be relatively less. And regardless of cost-efficiency, the interesting fact is that speakers of human languages must be capable of going beyond storage - whether they in fact use it 98 per cent of the time is beside the point.

## The emergence of structure in the acquisition process

With the view advocated above on the status of structure, there are some views on the way structure shows itself in acquisition that follow naturally. The learning process can be permitted to depend as much as possible on simple and processually manageable lexical-like process of recognition and retrieval of items. Structural features of the learning process are a matter of taking an extra step from simple recycling to reprocessing — just as structure in language comes into its own as an extra level above the purely lexical. Any step of that kind, on or off the learning target, is a step in the right direction because it gives the selection pressure something to operate on.

As emphasized by Melissa Bowerman, one of the best ways of recognizing structural learning at work in the acquisition process is the occurrence of persistent errors in combination patterns. The Danish child that said *sko op!* ('shoe up') for a while to mean that he wanted the shoes to be either put on or taken off, had clearly learnt that particles could indicate a transition of some kind. This interpretation could not plausibly be ascribed to a pure recognition/retrieval programme, since the collocation *sko op* is not likely to occur except purely by chance; and the element of differentiation of collaborating roles between the 'subject' component of the utterance and the 'transition' component is therefore most plausibly an example of the extra level of linguistic competence, which the child is trying his hand at here.

Tomasello and his collaborators (cf. e.g. 1996, 1997), have shown that there is good reason to assume that children proceed very cautiously from lexical-type learning to the combination patterns, and that the constructional approach is well suited to describe the learning path: abstract slots only very gradually emerge from the learning of alternative filler-combinations. This shows that abstract syntactic patterns are unlikely to be the stuff of which language learning is made. However, it does not show that language remains lexical in nature, once the step towards mastering syntactic relations has been taken. Emergence is a fact about the ontological history of complex phenomena - but once a higher-level phenomenon has emerged, as it were, it can no longer be subsumed under the lower-level foundations from which it arose.

## Final remarks

The point of these reflections is to argue for a type of complexity in understanding language and language learning that combines features of familiar, opposite points of view. The position I defend is essentially within the cognitive-functional approach as opposed to the formal, generative approach. Yet it emphasizes the place of syntax as a separate level of description, and of learner achievement. Rather than seeing syntax as autonomous, it agrees with functionalist assumptions in seeing syntax as grounded in non-syntactic features of language. But it also insists that syntax, though grounded in lexical properties and collaborating with them in all sorts of ways, also in some respects rises above them.

*English Department, University of Copenhagen, DK-2300 Njalsgade 84,*
*Copenhagen, Denmark, harder@coco.ihi.ku.dk*

## References

Bates, Elizabeth, Inge Bretherton & Lynn Snyder. 1988. From First Words to Grammar. Individual differences and dissociable mechanisms. Cambridge: CUP.

Bates, Elizabeth & Brian McWhinney. 1987. "Competition, variation, and language learning", in McWhinney (ed.)., 157-193.

Bates, Elizabeth, Donna Thal and Jeri S. Janowsky. 1992. Early Language Development and its neural correlates, ch. 5 in S.J.Segalowitz and I.Rapin (eds.) Handbook of Neuropsychology, Vol 7: Child Neuropsychology. (Series Editors: F.Boller and J. Grafman).

Bates, Elizabeth, P.S.Dale & Donna Thal.1994. "Individual differences and their implications for theories of language development", in P.Fletcher & Brian McWhinney (eds.), Handbook of Child Language.

Bowerman, Melissa. 1996. The origins of children's spatial semantic categories: cognitive versus linguistic determinants, in Gumperz & Levinson (eds.), 145-176.

Edelman, Gerald M.1992. Bright Air, Brilliant Fire. On the Matter of the Mind. New York: Basic Books.

Elman, Jeffrey L, Elizabeth A. Bates, Mark H. Johnson, Anette Karmiloff-Smith, Domenico Parisi & Kim Plunkett. 1996. Rethinking Innateness. Cambridge, MA: MIT Press.

Felix, Sascha W. 1987. Cognition and Language Growth. Dordrecht: Foris.

Fillmore, Charles J.—Paul Kay—Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: the case of *Let alone*, Language 64,3: 501–538.

Fodor, Jerry A. 1985. Precis of 'The Modularity of Mind'. Behavioral and Brain Sciences 8, 1-42.

Gumperz, John J. and Stephen C. Levinson. 1996. Rethinking linguistic relativity.(=Studies in the social and cultural foundations of language 17). Cambridge: CUP.

Greenfield, Patricia M. 1991. Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior, Behavioral and Brain Sciences 14: 531–595.

Locke, John L. 1996. Why do infants begin to talk? Language as an unintended consequence. Journal of Child Language 23, 2, 251-268.

Lund, Karen. 1996. Communicative function and language-specific structure in second language acquisition. A discussion of natural sequences of acquisition, in Engberg-Pedersen, Elisabeth—Michael Fortescue—Peter Harder—Lars Heltoft—Lisbeth Falster Jakobsen (eds.)

Content, Expression and Structure: Studies in Danish Functional Grammar. Amsterdam: John Benjamins, 385-420.

MacWhinney, Brian (ed.). 1987. Mechanisms of Language Acquisition. Hillsdale, N.J.: Lawrence Erlbaum.

Tomasello, Michael, N. Akhtar, K.Dodson and L. Rekau. 1996. Differential productivity in young children's use of nouns and verbs. Journal of Child Language 24: 373-387.

Tomasello, Michael and Nameera Akhtar. 1997. Young Children's Productivity With Word Order and Verb Morphology. Developmental Psychology.

# Placement and Scope of Mainland Scandinavian Modal Adverbs[1]

*Henrik Jørgensen*

## Abstract

This paper investigates the interplay between modal adverbs and other parts of the sentence, especially the grammatical subject, in the Mainland Scandinavian Languages, especially Danish. Two adverb positions are isolated and analyzed with respect to their status in sentence schemes (see examples 1-4 below). One of them is recognized as a normal sentence slot, whereas the other one must be treated as an adjunct to heads, rather than as an independent slot.

## 1. Introduction

### *1.1. The topic of the paper*

The topic of this paper is an investigation into the interplay between sentence adverbs and other parts in the sentence, notably the grammatical subject, in Danish and to some extent in the other Mainland Scandinavian languages. My main concern is two adverb positions not hitherto treated by the Diderichsen system of field syntax. As I shall attempt to show in my paper, both these positions are concerned with irregular focus positions in the discourse structure. This influences their empirical whereabouts and imposes stylistical and pragmatic limitations to the possibility of showing their existence. It also raises the question of whether a scheme with fixed maximal positions is actually the most efficient way to investigate Danish syntax. A central problem is the necessity to account for discursive features which influence the syntax. The Diderichsen scheme is conceived in such a way that a number of important observations are readily accounted for within the scheme, but the observations that I am dealing with here seem to resist this kind of "thinking in slots". I shall return to these speculations at the end of my paper.

### *1.2. Some preliminaries*

It is not strictly true that the two adverb positions dealt with here were 'not hitherto treated by the Diderichsen system of field syntax' as I suggested above. In fact, they have both been noted by Diderichsen, but never properly integrated into his scheme. But there is a methodological reservation about all this, too, and this pertains to the two problems: WHAT should be included in a sentence scheme and HOW to include it.

The traditional scheme looks like this[2]:

**Figure 1:**

| Forbin-derfelt | Fundament-felt (> Forfelt) | Neksusfelt (> Centralfelt) | | | Indholdsfelt (> Slutfelt) | | |
|---|---|---|---|---|---|---|---|
| **k** | **F** | **v** | **n (subj.)** | **a 1, 2...** | **V** | **N (obj 1-2)** | **A 1, 2 ...** |
| *og* | *saa* | *kunde* | *han* | *sikkert*<br>*ikke*<br>*alligevel* | *faa sagt* | *hende*<br>*Besked* | *i Tide* |
| 'and' | 'then' | 'could' | 'he | 'surely not anyway' | 'have said' | 'her a word' | 'soon enough' |

When Diderichsen started organizing his scheme in the late 1930s, he planned to organize it on a basis which is strongly reminiscent of modern dependency grammar, including notions of verbal valency etc.[3] However, in the final version dating from 1946, the scheme which forms the basis of the current discussion, he replaced it with a more down-to-earth principle, distinguishing elements of the sentence on a morphological basis. First, he distinguishes the verb, taking into account its split position between finite and infinite position. Next, he distinguishes elements of the sentence with a nominal kernel and a nominal function. Everything else is supposed to be adverbs; Diderichsen did reach some interesting new classifications of adverbs on the basis of their linear possibilities, but this was not foreseen from the beginning. In spite of this rather primitive and reductionist way of organizing the scheme, it is nevertheless able to show some interesting relational links, such as the subject and object positions, the object-like character of the situationally unknown subject, the treatment of indirect objects and their semantic equivalents. Also certain discursive functions find interesting expressions in the scheme, notably in connection with front position, which Diderichsen in his various accounts of Danish syntax has described as an important discursive factor.

It is also important in this respect that the scheme on this first level only contains slots[4] that are directly dependent on the verb. Elements of the sentence dependent on other elements below verb level have no slot in this scheme; they are accounted for according to their dependence on substantives (in the so-called 'genstandshelhed'), on adjectives ('beskriverhelhed') and adverbs ('adverbialhelhed'). This distinction between dependency levels of syntax is important, since it is an important factor when it comes to determining permutability.

Still, the question of what a slot in such a scheme represents remains open. I would suggest that a slot in the sentence scheme proper represents a discursively independent node in the dependency network directly dependent on the verb. This means that such a node is not represented by unstressed anaphors or by oversized constructions, since these last most often are subject to extraposition. A slot, then, has to contain at least one stress; and vice versa, that which has a stress and is a node directly dependent on the verb, should be admitted as having a distinct slot within the scheme to stand on. I have used a similar criterion (in Jørgensen 1991) to rule out the possibility that the landing site of the unstressed anaphors is a slot in any sense of the word. The criterion here is a refinement of similar, but less precise remarks in Jørgensen 1996.

There are three important reservations to be made in relation to this definition of 'slot'.

The first reservation pertains to the verb itself and is due to specific factors concerning the split verbal slot in Mainland Scandinavian, which is kept together as a syntactic and semantic entity by having only one stress:

* when an infinite verb is present, the finite verb has no stress, hence the v slot may be unstressed;
* in constructions with a verbal particle, the stress is on the particle in Danish and in many spoken versions of Swedish and Norwegian; hence the **V** slot may also be unstressed.[5]

The consequence is that this definition of 'slot' is only relevant to non-verbal elements of the sentence. The unificational stress of the discontinuous verbal chain through the sentence, uniting at most three different slots, operates with a different slot concept. Since this concept only works within the central dependency node, the definition of slot as given above is not affected in any crucial way.

The second reservation concerns with many sentence adverbs; adverbs like *nu*, *da*, *jo* and *sgu* (so-called modals) are inherently unstressed,[6] and hence would fall beyond the stress criterion suggested here. Nevertheless, in order to account for the kind of regularity that can be observed around such words, they will have to have a slot of their own. The most important argument is that they belong to paradigms containing adverbial constructions with an inherent stress and important semantic features closely related to those of the nonstresed adverbs mentioned above. Thus, the non-stressed character of these adverbs must count as an anomaly outside the scope of the greater lines of prosodic syntax.

The third reservation deals with the front position (forfelt), which may equally well be unstressed when filled with an anaphoric pronoun. Since I have claimed that such pronouns are clitic when they follow a stressed element of the sentence, it is possible to make a parallel claim concerning the front position, and suppose that such pronouns are proclitic. This, then, would be an instance of a general rule saying that slots determined by the valency of the verb may be filled with clitics following their own cliticization rules.

My conclusion is thus that these three reservations modify the concept of slots, but they do not fundamentally alter it.

### *1.3. Thesis*

The two slots to be discussed here are the following:

a. an adverb slot between the finite verb and the subject slot, as illustrated in these examples:

(1) Først vil altså maskinerne fylde op, og når så pludselig (...)[7]
‘At first the machines will fill up, and when then suddenly (...)’

(2) I Rumle Hammerichs ‘regeringstid’ har således 39 instruktører og 46 forfattere været involveret i produktionen i afdelingen, (...) {Politiken sect. II, Aug. 19th 1998.}
‘During the ‘reign’ of Rumle Hammerich have thus 39 directors and 46 authors been involved in the production of the department’

A traditional scheme is unable to account for this until an extra **a** slot is added:

**Figure 2:**

| **F** | **v** | **extra a** | **n** | **a** | **V** | **N** | **A** |
|---|---|---|---|---|---|---|---|
| Først | vil | altså | maskinerne | - | fylde | - | op |
| I Rumle Hammerichs ‘regeringstid’ | har | således | 39 instruktører og 64 forfattere | - | været involveret | - | i produktionen i afdelingen (...) |

The issue here is whether other solutions can be found, or whether there is good reason to extend the scheme, in spite of doubst one could have. My thesis is that such good reasons do exist and hence that such a slot has to be recognized.

b. Adverbs in the front position (‘forfelt’) together with subjects, other adverbs or other kinds of directly dependent sentence members, as in:

(3) Først kelterne lærte romerne at gå i bukser. {constr.}
'First the Celts taught the Romans to wear trousers.'

(4) Især Budapest er en spændende by. {constr.}
'Especially Budapest is a fascinating city.'

Diderichsen (1962 §76-2 mentioned the existence of these adverbs, and also modified the conventional dogma on the front position being able to contain only one element of the sentence in order to account for these examples (Diderichsen 1962 p. 190). My task here will be to investigate the conditions of this interesting exemption from the principles of the front position and at the same time to discuss whether there are any implications for the analysis of such sentences. Unlike the first subject of discussion (point a. above), my claim will be that separate slots are not needed here.

## 2. Sentence adverbs before the (inverted) sentence subject

In spite of the fact that this slot never found a place in the elementary University grammar books on Danish, it has nevertheless been noted by many observers[8]. Lars Heltoft (1989 p. 141) has fully integrated it into his recent revisions of the sentence scheme. My own previous treatment of the subject (Jørgensen 1996) was an attempt to clarify stylistic and empirical issues in relation to this slot. In this discussion, I want to focus on its prosodic and discursive functions .

The Swedish and Norwegian presentations of this slot are more comprehensive, since it seems to play a more important role in these two languages.[9] One reason for this is the fact that you often find the negation in this position in Swedish and Norwegian, contrary to Standard Danish, where the negation occurs only marginally in this position. However, it is possible to argue that the Swedish and Norwegian negation - contrary to the Standard Danish - is enclitic to the **v** slot. This can be observed from its phonetic form alone, and hence raises the question of whether the behavior of the Swedish and Norwegian negation is not better accounted for through syntactic cliticization.

The same question could be raised in connection with the Danish examples, since many of the adverbs found in the actual examples are unstressed. However, enough examples with obviously stressed filling of the slot remain to allow us to claim that the slot is actually relevant to the criterion:

(5) Men det måtte naturligvis tiden vise. {government paper}
'But this had of course time to show.'

(6) Mangler Subjektet, eller vilde dets Foranstilling medføre en falsk Pointering, kan undertiden Sætningsadverbialet staa paa Indlederplads, vistnok uden særlig Emfase. {Diderichsen}
'If the subject is missing, or if its front position would induce a false emphasis, can sometimes the sentence adverb stand in front position, possibly without particular weight.'

All these examples are obvious cases of a proper subject in the subject slot, immediately preceded by a stressed adverbial construction. If we allow the unstressed, non-clitic adverbs to use the same slot, there is a good deal of evidence to suggest that there is actually such a slot satisfying the criteria set up in the introduction.

It is important to note, though, that the examples found with this slot filled out are usually quite simple in their construction. Usually the **V** slot is empty and quite often the verb is also intransitive so the **N** slot as well is, therefore, typically empty. In general there is a strong tendency for the Danish examples to show up in sentences whose semantic content would designate subjects being in certain places, showing up or disappearing from certain contexts. Thus they often constitute a semantic parallel to existential sentences where the (semantic) subject is also the focused element. However there is a slight difference here since the subject in the adverb-before-subject construction is very often a name or a well-known entity in some other sense. To state the existence of such an entity with an existential sentence proper seems inappropriate: since the existence of the subject is well known, the only new effect is to put it into a new context (which may be the center of the focus in the actual sentence.

Although similar from a formal point of view, a special group of examples seems to call for a different explanation, namely a-slots before subjects in wh-questions:

(7) Men hvor kommer egentlig dette dyr fra?
'But where comes in fact this animal from?'

(8) (...) hvordan skulle så politikerne i et lille frynseland som Danmark kunne bestemme noget som helst?
'(...) how should then the politicians in a small fringe country like Denmark be able to decide anything at all?'

(9) Hvis Robert Healey havde druknet sig, hvor var så hans kone og steddatter?
'If Robert Healey had drowned himself, where were then his wife and stepdaughter?'

Even the negation may be found in questions, in spite of the fact that this is otherwise not possible:

(10) Kan ikke dette forhold indvirke på løsningen? [constr.]
'Can not this matter influence [on] the solution?'

If the focus in a wh-question is the question marker, then the adverb obviously has a different function here, namely to point to the subject of the sentence as an important presuppositional factor to the answer. I suppose that this is the basis for these examples, but I have not yet investigated this matter in any detail.

## 3. Incorporated sentence adverbs

My attention was drawn to this group while investigating the independent sentence adverbs. Occasionally one would find examples looking very much like independent adverbs in front of sentence subjects, which would then have to be discarded because they would move with the sentence subject when this was moved e.g. to front position (forfelt). Since it is a useful dogma that the front position can contain only one element of the sentence belonging to the primary level, such examples were obviously irrelevant to the discussion raised in section 2.

Nevertheless, these adverbs presented a difficulty in yet another sense, since they were obviously found in positions where they acted as modifiers to NPs and in fact were integrated into NPs, obviously being permutable with the NP. This is not easily brought into harmony with the traditional notion of adverbs being modifiers to verbs or to whole sentences. Of course no one would deduce weighty conclusions from traditional grammatical dogmas, but the clash between the observation and the traditional theory in this case is strong enough to arouse suspicion.

The Diderichsen tradition has done very little work on structures below the primary level of classical sentence elements, and the fact that certain adverbs usually regarded as 'sentence adverbs', may appear in these positions has never been closely investigated. Diderichsen has noted the phenomenon in § 65-1 and § 76-2, both times only briefly stating facts without entering into any kind of deeper description. In his description of NPs ('nominalhelheder'), he does not acknowledge the possibility of having adverbs in front of such constructions; therefore, the two short remarks are obviously intended to suspend the dogma of 'only one phrase in front position'. This is, admittedly, only one possible interpretation, but on the other hand, the joint permutability remains as an obscuring fact. One short remark from Diderichsen seems to be an important clue: he states that the adverbs in this construction are unstressed. This is proved by the examples I have given, and there is therefore an important point of syntactic prosody here. I shall return to it later.

A brainstorm[10] yielded the following list of Danish adverbs and adverb phrases; it is expanded here with examples from a Danish text corpus illustrating their function in relation to subjects, time- and place-adverbs and other types of sentence-elements:

**Figure 3:**

- akkurat
- alene
- allerede
- blandt andet/andre (bl. a)
- heller ikke
- i det mindste
- i høj grad
- ikke alene
- lige præcis
- navnlig
- netop
- nøjagtig

- blot
- derimod
- endnu
- ene og alene
- f. eks.
- frem for nogen/noget
- frem for alt
- først og fremmest
- først
- ikke kun
- ikke engang
- ikke mindste
- ikke så meget som
- indtil videre
- især
- kun
- lige
-lige netop
- også
- præcis
- selv
- sidst men ikke mindst
- specielt
- sågar
- under alle omstændigheder

Actually, the variety of sentence-elements to which such adverbs could be attached is quite large. This is illustrated here by 'alene':

*- subject:*

(11) (...) og da byrådet har besluttet at afskaffe ventelisterne over et kort åremål, vil alene denne konto kræve store investeringer i anlæg og drift fremover.
'... and since the town council has decided to get rid of the waiting lists over a short span of time, will alone this account demand great investments in material and service in the years to come.'

*- subject with adverb attached*

(12) Eksporten alene er næsten 190 milliarder kroner, (...)
'The export alone is almost 190 bill. kr., (...)'

- *different scenic adverbs:*

(13) Alene i London var flere hundrede veje spærret af væltede træer.
'Alone in London were several hundred roads blocked by fallen trees.'
(14) Alene siden påske sidste år er 735 mennesker dræbt og over 12.000 kvæstet ved trafikulykker.
'Alone since Easter last year have 735 people been killed and more than 12,000 injured in traffic accidents.'

The crux of these syntactic phenomena is that the adverbs obviously retain their normal semantic content while at the same time functioning below the normal element-of-sentence level. In fact they may show up at any border line in the syntax and start the scope of their modification from almost any break:

(15) (...) for hvordan skulle de høje herrer på slottet dernede ellers have fundet frem til akkurat ham.
'(...) since how should the sires in the castle down there otherwise have found forth to precisely him.'
(16) (...) en del af denne forøgelse skyldes dog opskrivning af ejendommene i København og Århus, som på blot to år er opvurderet fra 16 til 26 mill. kr.
'(...) a part of this growth is due, however, to the reassesment of the houses in Copenhagen and Århus, that in only two years have been written up from 16 to 26 mio. kr.'
(17) (...) han gør det også muligt for publikum at komme ind i en situation med bare et par replikker.

'(...) he makes it also possible for the audience to enter into a situation with only a couple of lines.

(18) Dvs. med et venligtsindet eller i det mindste neutralt styre [i] Kabul, (...)

'That is, with a friendly or at least neutral government in Kabul (...)

It is worth noting that the scope of these adverbs is limited to the constituent in which they occur; e.g. the limitation of quantity in *blot* in (16) pertains only to 'two', but certainly not to the figures later in the same sentence. In such cases, a change or a levelling of the meaning would normally be expected, but quite clearly such a change or levelling does not occur. However this is not quite as odd as one might think. There seems to be little reason to assume regular slots for such adverb positions; the most sensible thing to do would be to add the possibility of adverbs entering as adjuncts in front of any slot on any level to exercise their scope from the point in question. It is worth noting that the technical solution here is foreign to any kind of thinking based on slots.; This is in keeping with the author's fundamental assumption concerning slot systems like Diderichsen's, i.e., that they have no independent existence, but are merely a convenient way of circumscribing more complicated dependency structures.

In fact, all these adverbs are normally considered to be sentence adverbs and they, therefore, belong to a group of words whose inclusion in the group of adverbs proper is, in any case, doubtful. Hans Götzsche (forthcoming) calls this group of adverbs 'quasi-adverbs', separating them from the traditional syntactic-semantic groups of adverbs. He points out that the semantic status of these quasi-adverbs is often that they reflect another sentence which acts as a presupposition (in a non-formal sense) to the actual statement being uttered. This would seem to be the semantic impact of these adverbs in all these constructions and, from a semantic point of view, the syntactic integration seems rather misleading.

Diderichsen seems to have opted against any theory of incorporation and, in fact, the behavior of such adverbs seems to escape traditional slot conceptions of syntax.

An important question is what kind of semantic impact these adverbs have. A majority of those that can be observed in these constructions obviously have a focusing sense, but a special one relating them to cleft sentences. They mark a contrastive focus, highlighting the focussed element against a group of elements relevant to the statement that is made, but not actually included in it. A statement like:

(19) Kun Olsen har set løsningen. {constr.}
'Only Olsen has seen the solution.'

assumes that the solution might have been found by several (unmentioned) individuals somehow engaged in the search - as opposed to another group outside this action, - but highlights Olsen as the one to whom the statement really applies. In fact this highlighting effect is present in all cases, even though it may show up in a somewhat tricky form, like this:

(20) Selv borgerskabets bedste børn er rendt hjemmefra, blot for at se på {C.V.Jørgensen}
'Even the upper ten's best children have left home, only to look on.'

In this case, the group nominated is a group of individuals to whom the relevance of the statement seems unlikely - as opposed to a group that would be prone to leave home in any case - and the element highlighted is the most unlikely member of the nominated group.

All these content structures are closely related to a semantic mechanism first described by Carl Vikner (1970) in connection with French cleft sentences. Vikner, in fact, uses the same kind of set theory description as I do. The fact that the semantic description is identical indicates that all these phenomena are instances of a much larger discursive structure, namely the array of means to modify focusing structure within a normal sentence string.

Contrasting focuses are complex discursive matters. They relate to statements to which the speaker cannot be said to have the full copyright, so to speak. Rather, these statements are 'in the air' and must be seen as echoes of statements (actually uttered or presupposed) by other speakers. This is why their linear order may require a modification of the focusing which interferes with the original organization of the semantic material, even down to tearing up structures knit below normal sentence level structure. Such considerations may explain why these structures have semantic qualities on the discourse level of the sentence, while at the same time having surface syntactic structures which set them on the same footing as elements below sentence level.

## 4. Concluding remarks

In the light of the conclusion to the last section one could ask a more general question: do sentence adverbs have slots at all? Do our results with the contrasting adverbs not suggest that no sentence adverb has a slot under any circumstance?

I think that the non-contrasting adverbs do have a slot, and the reason is that their linear behavior at normal sentence level is regular enough to justify this slot. 'Slotlessness' is only a characteristic of the contrasting adverbs due to their specific scope of operation, simply because their scope may highlight very specific sections of a sentence string. The adverbs in non-contrasting sentences have more conventional effects, and hence also slots to exercise them from.

The most important point to be made is that slot systems, like the sentence scheme - in spite of all the reservations one can have about the generality of such an apparatus - may have their virtues in actually forcing the syntactician to reflect semantic, discursive and pragmatic problems at the same time as he is trying to piece his slot machine together. A slot system per se is not what I would like to put into force; I think rather that the slot system should be organized in such a way that the discursive considerations immediately force themselves upon the reader of the textbook.

The main task, in my opinion, seems to be incorporating the topic-and-focus structure into the sentence scheme. In itself, topic-and-focus is an important syntactic parameter, which the Diderichsen tradition has noticed as well, especially in connection with the front position. The problem rises when other focusing structures occur, such as cleft sentences or the focusing adverbs treated here in section 3. Such structures quite often break the linear order imposed by the normal sentence scheme. The best way to account for such regularities seems to be to introduce in a very simple way a parallel linearity underlying the sentence scheme, joining its elements on a line. In the simplest case, this line moves from topic to focus and, under conditions to be specified as precisely as possible, it may be modified to suit the discursive means.

The ultimate goal of this line of re-structuring the sentence scheme is to see the slots recognized in the Diderichsen tradition as an output of exact semantic and discursive factors yet to be described. But this is a vision, not a fact, and I think that this little glimpse of a Danish syntax to come is a suitable place to stop.

*Institut for nordisk sprog og litteratur, Århus Universitet*
*Niels Juelsgade 84, DK-8200 Århus N, Denmark*

## References

Basbøll, Hans 1986: Diderichsen vs. Dik eller Feltanalyse vs. Funktionel Grammatik. Nydanske Studier 16-17. 56 - 76.

Bleken, Brynjulf 1971: Om setningsskjemaet. Oslo - Bergen - Tromsø: Universitetsforlaget.

Bruås, Einar 1971: Eksperimentell-analytisk lingvistikk. Oslo - Bergen - Tromsø: Universitetsforlaget

Diderichsen, Paul 1936: Prolegomena til en metodisk dansk Syntax. Repr. in Helhed og struktur. København: Gads Forlag 1966. 21 - 24.

Diderichsen, Paul 1962: Elementær dansk grammatik. 3rd ed. København: Gyldendal.

Diderichsen, Paul 1964: Sætningsleddene og deres stilling - tredive år efter. Repr. in Helhed og struktur. København: Gads Forlag 1966. 364 - 379.

Götzsche, Hans (forthcoming): Om satsadverbial i svenska. To appear in Svenskans beskrivning 23.

Hansen, Erik 1970: Sætningsskema og verbalskemaer. Nydanske studier 2. 115 - 142.

Heltoft, Lars 1986: Topologi og syntaks. Nydanske studier 16-17. 105 - 130.

Heltoft, Lars 1989: Talesprog og sætningsskema. 2. Møde om Udforskningen af Dansk Sprog (ed. Kunøe & Vive Larsen). Århus 1989. 128 - 152.

Heltoft, Lars 1992: Topologiens plads i en sprogteori. Sprogvidenskabelige arbejdspapirer fra Københavns Universitet. Årg. 2/ 1992. 67 - 98.

Jørgensen, Henrik 1991: Om de personlige danske pronominer. Danske studier 1991. 5 - 28.

Jørgensen, Henrik 1996: Om adverbialled mellem det finitte verbal og subjektspladsen i danske hovedsætninger. Selskab for Nordisk Filologi København. Årsberetning 1994-1995 (ed. Henrik Galberg Jacobsen). 76 - 90.

Jørgensen, Henrik (forthcoming): Studien zur Morphologie und Syntax der Festlandskandinavischen Personalpronomina. To appear at The Department of Scandinavian Studies, University of Århus.

Thorell, Olof 1982: Svensk grammatik. 2nd ed. Stockholm. Esselte.

Vikner, Carl 1970: Quelques réflections sur les phrases clivées en français moderne. = RIDS (Romansk Instituts Duplikerede Småskrifter) 12. Mimeo, Københavns Universitet, Romansk Institut 1970.

Vive Larsen, Erik 1986: Paul Diderichsens sætningsskema og den nordiske syntaksforskning.Nogle centrale problemstillinger. Nydanske studier 16-17. 131 - 148.

## Notes

---

1. I am grateful to Sten Vikner for comments on the final version of this paper and to Patricia Lunddahl, Faculty of Humanities, Translator Service, Århus, for her revision of the English text. Of course, any remaining inaccuracies are my own responsibility.

2. Diderichsen 1962, p. 186, slightly modified and with the present field terminology added, cp. Diderichsen 1964, Hansen 1984.

3. Diderichsen 1936 (=1966 p. 21-24), and Jørgensen (forthcoming) chap. 1

4. *Slot* is used in this paper as a translation of the Diderichsen term *plads*. In other English presentations the term *place* is used as a translation, but this term is used here in a non-technical sense, as opposed to the technical definition of slot.

5. Since the particles in Swedish and present-day Norwegian are between the infinite verbs and normally-stressed objects ('ta på mössen', 'kjøre ut bilen'), it is an open question whether the particles are actually in a **V** slot or in an independent **P** slot, such as the one suggested for Danish by Heltoft (1992 p. 75 et passim).

6. Of course the homonyms belonging to other word-classes do not apply in connection with this statement.

7. Unless otherwise stated (in { }), examples are from Henning Bergenholz' corpus DK87-90.

8. Hansen 1970 p. 132; Basbøll 1986 p. 71; Vive Larsen 1986 p. 142, Heltoft 1986 p. 129.

9. Bruås 1971 s. 54, Bleken 1971 s. 50f, Thoréll 2nd. ed. 1982 § 746

10. I am grateful to Birgitte Skovby Rasmussen for her many contributions to this brainstorm and also for her suggesting many of the points later brought up in connection with the semantic interpretation of the adverbs in the list and the construction of which they are a part.

# Om begreppet kognitiv situation

*Oili Karihalme*

**Abstract**

In linguistic semantics, it is difficult to take into consideration the wholeness of knowledge of a person or a group. The research of special languages in particular remains formalistic and sometimes superficial, because the scale of knowledge is not understood widely enough. If we adopt the concept of *cognitive situation*, in which we collect intension from several branches of research, especially philosophy, action science, and mental capasity research, we are able to understand better the semantic grounds of utterances of language. The research will reach from the nonverbal to complex linguistic situations. I concentrate in my present survey on cognitive situations in which the possibilities to use verbal utterances are in some respects restricted. The linguistic practices of my special language testgroups offer promising material for the developing process of the idea of cognitive situation.

Kognitivisterna frågar sig om och om igen, med vilka metoder man kunde fånga kunskaps-helheten (*the total knowledge state*) hos en individ eller en viss grupp, om det överhuvudtaget är möjligt. Hur kunde man i forskningen förena den teoretiska och praktiska kunskapen med människans kognitiva kapaciteter?[1] En kognitiv semantiker måste ju vara medveten om allt som påverkar de semantiska innehållen och relationerna i handlande, tänkande och språk.[2]

När man granskar språkets relation till de mänskliga kunskapsresurserna vid olika tillfällen är det lätt att konstatera, hur problematiskt det kan bli att göra sig en föreställning om kunskapssammansättningen i en enskild språkbrukssituation, till exempel när någon använder en specialiserad term. Än så länge har man inte lyckats binda samman den kognitiva språkforskningen sådan som den nu är med synpunkter som hör samman med fack- eller specialkunskaper och fackspråklig kommunikation. Den leksikala semantiken av fackspråk har kanske därför förblivit formalistisk och ofta också ytlig. På 80-talet begränsade bland andra de så kallade *minimalistsemantikerna*[3] utvecklingen av den lingvistiska forskningsbranschen, som skulle basera sig på fackkunskapens semantik. De ansåg att alla betydelsefenomen inom fackspråken är operationer som enbart hör till den separata begreppsliga nivån. Man måste därför hålla dem avskilda från den rent lingvistiska lexikala semantiken. Det är emellertid mycket nyttigt och intressant för en lingvist att granska olikheterna mellan den allmänna och den speciella kunskapsbakgrunden. Arbetet hjälper oss att förstå på ett djupare sätt det kognitivas andel i språket (och vice versa).[4]

När fackpråken inryms i den språksemantiska helheten, förutsätts lingvisten skaffa sig en djupgående kännedom om specialområdet och dess kunskap. Man måste helst känna till sätten för specialkunskapens uppkomst och specialisternas sätt att sätta sina mål, kategorisera, eller icke kategorisera, och ge uttryck åt kunskap. Själv satte jag mig in i formgivningens kunskapsteoretiska bakgrund medan jag skrev min doktorsavhandling Specialiseringen av formgivningens teoriord [Muotoilun teoriasanaston termistyminen] (1996). Formgivarnas yrke är ett särskilt fint exempel på en bransch, där diskursen i sin helhet blir oförstådd, om man inte har tålamod att reda ut, vilka typer av kunskap och teorier som framträder på området och speciellt vad som kan uttryckas språkligt. Språkets semantik har nämligen i detta fall sin grund både i reflexionen och den praktiska, mångfaldiga arbetsprocessen. Alla olika produkter av designprocessen, alltså objekt för projektering och tillverkning bär mycket varierande kulturbunden men annars ospecialiserad kunskap och ytterligare en överraskande mängd av specialiserad kunskap i sina så kallade *betydelsekonstruktioner* som vanligen helt öppnar sig bara för de yrkesverksamma. Formgivningens snabba begreppsliga specialiseringsprocess får mest stöd av internationella teorier, som gemensamt kallas *produktsemantik*.[5] Alla vardagliga, tekniska eller estetiska föremål, till exempel *hiss*, *sits* eller *UPS* kan få en mycket specialiserad semantisk, språklig beskrivning genom dessa teorier.

Till kunskapskartläggningen av en specialbransch hör att utröna och definiera de kognitiva situationer som de inom branschen verksamma personerna eller grupperna befinner sig i. Jag binder begreppet *kognitiv situation* med språket och definierar det här preliminärt som den situation som råder när en grupp eller individ skaffar, bemästrar och / eller manipulerar kunskap och som direkt eller indirekt påverkar språket och språkbruket. Begreppet kunde man metodiskt närma sig också via den så kallade *situationssemantiken*[6] , i synnerhet om företrädarna för denna semantiska riktning själva skulle vara klarare i fråga om sina grundbegrepp, deras innehåll och brukbarhet. Inom situationssemantiken använder man begreppet *resurssituation*. Begreppet kunde möjligen ge stöd för en definition av de kognitiva situationer som förekommer inom vetenskap, fackområden och andra specialiseringar. Tyvärr har man tillsvidare inte kunnat nå överenskommelse om en så fundamental detalj, som om man med *resurssituation* avser den totala kunskapsgrunden som i allmänhet finns till i en viss språkbrukssituation eller bara kombinationen av de textbundna kunskaper, som alltså kan plockas direkt ur texten. Detta visar, hur man kontinuerligt vacklar mellan den traditionellt filosofiska och den traditionellt språkliga synvinkeln. Det verkar svårt att sammansmälta de båda.

Ett ytterligare hinder för utnyttjandet av begreppet *resurssituation* (Cooper 1996: 76B77) är att diskussionen om situationens ontologiska natur, bland annat dess koherens och intension, ännu inte har lett till några tillfredsställande slutsatser. Åsikterna går häftigt isär. Är situationen alltså en sammansättning av några tydligt skilda drag eller fenomen i kontexten (är den ett *set*)? Kan man tillfredställande förteckna och beskriva dessa drag med en vetenskaplig, semantisk metod? Frågan ligger nära kategoriseringsproblematiken. En finsk forskare inom formgivningen, Keinonen (1992: 60B63), har sett den semantiska informationssituationen i design som en *black box*, som man måste kunna dela i bitar om man vill dra någon nytta av den.

De semantiker, som understöder bruket av situationsbegreppet, hänvisar ofta till att en sammansättning av separata situationsdrag kan man förstå endast på en mycket hög abstraktionsnivå. Däremot kommer det **holistiska, odelade situationsbegreppet** mycket nära den reella, dels reflekterande och dels omedvetna processen, genom vilken en viss agent grupperar situationsinnehållet på sitt eget individuella sätt. När det är fråga om fackspråk, kan både en grupp av specialister och en enskild fackutbildad individ anses som en sådan **kunskapsagent**. Det kan anses vara motiverat att tänka sig att alla individuella och också samhälleliga reflexions- och kunskapsprocesser är i sig själva så betydande att det är berättigat att tala om *kunskapssituationer* eller *kognitiva situationer* i den mening som situationssemantiken gör. Från språkforskningens synpunkt är den bästa motiveringen för begreppet *kognitiv situation* att språkbrukarens eller språkbrukarnas kunskap om språket tillsammans med deras övriga kunskap eller sinnesinnehåll alltid på något verifierbart sätt återspeglas i deras språkliga uttalanden. Detta påstående upprepar den såkallade *i-semantikens* (*internalized semantics*) grundprincip. Man ser all betydelse i språket i förhållande till världsbilden (*grasp / understanding of the world*) hos språkanvändaren.[7]

Den kanske möjliga hypotesen att det skulle vara enkelt att utreda sinnesinnehållet hos specialisterna därför att deras kunskap är så exakt, ramlar på de första meterna. All specialkunskap är inte nödvändigtvis propositionell eller logisk. Också de diakroniska och andra dynamiska processer som hela tiden pågår i kunskapshelheter och i språkbruk försvårar situationsforskningen. Situationerna är aldrig precis likadana, i verkligheten är det så att varje gång ett språkligt uttryck används är situationen ny, unik. Semantikerna har alltid varit medvetna om att de inte förmår observera eller beskriva alla dynamiska processer, till exempel utvecklingen av diskurstyp och diskursspråk, tillägnandet av diskurssätt samt alla språkbrukets reflekterings- och korrigeringsprocesser till vilka man också kan räkna språkvården.

Jag koncentrerar mig här på ett grundläggande problem som angår den kognitiva språksituationen. Frågan har varit viktig för min fortsatta undersökning av formgivarspråket. Den gäller gränsdragningen mellan de ordlösa och verbala kognitiva situationerna. I språkforskningen måste man alltid minnas hur viktigt det ordlösa är för det verbala. På specialbranscher uppstår det mycket komplicerade system av arbetsfördelning mellan den praktiska handlingen och språket. Detta påverkar alla språkliga nivåer, textnivån såsom termerna.

Bland formgivarna finns det många personer vilkas attityder ger oss ett gott exempel på hur den ordlösa yrkesverksamhetens starka ställning kan påverka den språkliga identiteten och därigenom också språket. Begreppsliga och språkliga kunskapsresurser, speciellt exakta termer, anses inte angelägna. Klyftan mellan praktikens övare och teorins älskare är djup. Kunskapsspråket förblir outvecklat eller är antaget bara av den reflekterande delen av yrkesutövarna.

För att kunna analysera situationerna i praktiken och systematisera deras förhållande till språket och språkbruket behöver man en helhetssyn på kunskapsnivåer som i allmänhet förekommer. Man kan se sex olika faser i övergången från den ickeverbala situationen till den verbala. Källorna för den följande enkla listan är många från filosofin och pedagogiken till handlingsforskningen och den kognitiva vetenskapen X men inte lingvistiken därför att frågan har inte varit populär inom den språkvetenskapliga forskningen.

1. **Handling grundar sig på imitation av praktiska exempel**. Exempel följer man själv-villkorligt eller för att man blir tillsagd att göra så. **Språk kan förekomma i dessa situationer bara sekundärt**. De handlande agenter har ingen avsikt att förklara eller beskriva några entiteter. Kunskapen i den aktuella situationen är tyst, men kan nog diskuteras till exempel av filosoferna utanför situationen.

2. **Det uppkommer generella handlingsregler**. I praktiken är reglerna mest implicita och således **tysta**. De avancerade utövarna av de praktiska färdigheterna är kompetenta att ta emot och följa också **språkliga anvisningar**.

3. **Handlingen blir uppenbart intentionell**. Det uppkommer värderingar och kvalitetskrav för handlingen och dess resultat. Man behöver **reflexion** för att kunna förbättra prestationer. Skickligheten blir alltför viktigare och man börjar forma handlingsregler och beskriva entiteter. Även situationer som avviker från det normala eller vanliga på något sätt, framkallar behov att yttra sig verbalt. Man har påstått, att själva reflexionen inte är i grunden biologisk eller psykologisk utan social. Begreppsligandet (begreppsskapandet) och minnandet däremot är hjärnbaserade kognitiva processer. Här börjar alltså det kognitiva i mera inskränkt bemärkelse.

230

4. **Handlingen blir beskriven i ord.** Den verbala beskrivningen är i detta skede oftast en situationsbunden berättelse. Man anser allmänt, att berättelsen är redan en sorts uttryck för ett begrepp. Med hjälp av berättelsen kan man **identifiera** en situation, en process eller ett annat sakläge om och om igen. Uppkomsten av begreppsbildande berättelser är starkt beroende av den sociala situationen så att ju öppnare och vänligare situationer är, desto gynsammare blir de för sådant primitivt begreppsskapande.

5. **Exakta verbala exempel tas i bruk**. Man skapar **hypoteser** och **teorier**, som kan vara vardagliga eller också på något sätt specialiserade och tankemässigt krävande. För att kunna framhäva teorier måste man vara kapabel att utnyttja propositioner och kategorier.

6. **Man flyttar över till det medvetna begreppsskapandet**. Man bildar och definierar begrepp och ger dem namn. Man är i någon mån medveten om att begreppen är dynamiska. Den tekniska och vetenskapliga utvecklingen, alltså utvecklingen på fackområdena, utsätter begreppen för kontinuerlig förändring. Förnyelsen av begreppen sker via sk påståendelänkar. Sinnesinnehållen, som är viktiga att känna till när man definierar den aktuella kognitiva situationen, uppstår så att reflexionbaserade påståenden knyter ihop ny kunskap med de gamla begreppsinnehållen som befinner sig i långtidsminnet. I något skede orsakar detta ändringar i språket. Termernas preciseringsprocess är det genomskinligaste fenomenet.

Inom handlingens område finns det enligt den ovan introducerade modellen den praktiska, tysta och dels också omedvetna handlingens delområde, men där måste också ingå >det epistemologiska momentet= som Janik (1996: 24) kallar det. Förutom de praktiska filosoferna har den pragmatiska men å andra sidan epistemologiska handlingen intresserat även de kognitiva forskarna på senaste tiden. Kirsch och Maglio (1994: 513) har gjort vissa försök att skilja åt det pragmatiska och det epistemiska i handlingen. Enligt deras definition är den pragmatiska handlingens intention att bringa agenten fysiskt närmare ändamålet. Den epistemiska handlingen däremot syftar till att lyfta fram sådan information, som ligger gömd eller är besvärlig att hantera mentalt. Information kan man blotta både med hjälp av språket och utan det, men det som är svårhanterligt mentalt är också svårt att verbalisera.

Begreppet *kognitiv situation* är sådant till sin natur att man inte enbart kan nöja sig med det teoretiska resonemanget. Därför har jag börjat det pågående lingvistiska forskningsprojektet med några experiment, som syftar till att samla språkligt material och därigenom konkret evidens för situationer, i vilka det av något skäl är mödosamt att uttrycka sig med ord. Senare skall jag återvända till formgivarnas arbetsspråk och granska gränsområdet mellan det verbala och ordlösa mera holistiskt. Ordlösheten har ju många skäl. Man använder inte maximalt de kunskapliga resurserna som man har. Man kanske inte vill ge kunskap något uttryck alls. Ibland tar man inte i bruk de språkliga och andra möjliga yttringarna som man redan känner till. Formgivarna tycker ganska allmänt att språket stör och begränsar den kreativa processen. Gemensamt för de ifrågavarande testsituationerna alltså var att försökspersonernas kunskapsresurser var otillräckliga.

Jag har försökt få grepp om de besvärliga kunskapssituationernas diskurssätt och språk genom att arrangera bland annat krävande gruppdiskussioner. Ämnena har varit på något sätt abstrakta, komplicerade, irriterande eller dåligt kända. Diskussionssituationerna har realiserats vid olika kommunikationskurser.Testpersonerna har varit studerande och lärare av formgivning vid yrkeshögskolan i Kuopio och ekonomistuderande vid Vasa universitet.

Femton grupper av ekonomistuderande (minst 4 och högst 10 personer i en grupp) har diskuterat formgivningens betydelse och påverkan på marknaderna. Medlemmarna i diskussionsgrupperna visste att jag samlade material för en språkvetenskaplig utredning. Kunskapssituationerna är alltid mycket känsliga: man kan inte förödmjuka deltagarna med att utsätta dem för intellektuella svårigheter utan att förklara vad syftet med allt är.

Om man tänker på de sex faserna, som jag presenterade ovan, kan man säga att universitetsstuderandenas värsta problem låg på den femte nivån. De förstod på förhand, vilka krav situationen ställde. De visste att man väntade skickligt begreppsbruk och klara definitioner av dem. De var väl bekanta med den akademiska tekniken för språklig abstraktion. Studenterna befann sig i en svår konfliktsituation mellan att veta och inte kunna:

*-- jos, jos todella, jokaisella on oma, omat tällaset mieli-, tuot-, tai sellaset, muotosa ja se ei välttämättä, mene kymmenessä vuodessa pois. Ja toisaalta se, kauneus luo viihtyvyyttä ja myös sitten parantaa sitä mielenterveyttä ja sitä henkistä puoltakin että...*

[--om, om verkligen, var och en har sin egen, egna sådana favorit-, prod-, eller så här, sina former och det går inte nödvändigtvis, bort i tio år. Och å ena sidan den, skönheten skapar trivsel och också då främjar mentalhygien och den andliga sidan att...]

Man kunde tänka sig att studenterna skulle ha valt den fjärde nivån för begreppsligandet och berättat anekdoter om sina vardagsföremål eller använt andra narrativa medel. Dem änvänder man nämligen när kunskapsnivån är låg (eller den praktiska handlingen på ett visst sätt dominerar). Narrativt språk var dock mycket sällsynt i studenternas diskurs. Egentliga berättelser framkom inte alls i diskussioner. Generisk användning av pronominet *du* som inte är mycket frekvent och inte heller rekommenderas i finskan, var ovanligt rikligt i bruk. Dess förekomst var ett klart kännetecken på narrativ generalisering. De få exemplen som diskussionsdeltagarna använde var enbart korta hänvisningar på några parallella fenomen i vardagslivet.

Det var antagligen abstrakheten eller vetenskapligheten hos det givna diskussionsämnet, som förhindrade användningen av narrativa lösningar. Försöksgrupperna var också heterogena, de bildade inte någon homogen arbetsgrupp, såsom de yrkesverksamma ofta gör. De flesta av studerandena hade inte tillräckligt mod att tala personligt, öppet och fritt. Å andra sidan var mina försökspersoner mest överraskande i det att de tydligt visade hur väl man kan behärska en situation och uttrycka tankemodeller med retoriska formuleringar i en diskussion, fastän man inte kan utnyttja relevanta termer. Det språkliga yttrandet varierade starkt individuellt vilket beror på att några av de medverkande bemästrade allmänna definitionsregler för begrepp bättre än de andra. De hade alltså fina resurser när det gällde kunskapssystematik. Mina upptäckter är parallella med det allmänna påståendet bland de praktiska filosoferna att om språket speglar modeller för kognitiva operationer är det lätt inlärt.[8] Omvänt kan man påstå att man kan lätt producera operativt språk också när man har dåliga kunskapsresurser.

Diskussionsämnet för formgivarna hette *Hur kan man identifiera en >formgivningens produkt=?* Försöksgrupperna, som sammansattes av lärare vid en yrkeshögskola hade svårigheter på den sjätte nivån. Till detta finns det många orsaker. En av dem är den snabba kognitiva och terminologiska utvecklingen inom fackområdet. Alla orkar inte hänga med. Dessutom formgivningsspråket har inte institutionaliserats i den takt som man kanske borde. Det finns till exempel inga fackordböcker för teoriord. En annan intressant orsak är den negativt laddade kunskapsidentiteten i branschen. Det är viktigt att kunna men inte så viktigt att analysera.

Den sociala aspekten hos den kognitiva situationen på den sjätte, reflektiva nivån skall inte försummas. Både i den filosofiska och i den pedagogiska litteraturen framhävs den reflexionens gemensamma, gruppsorienterade natur. *Shared knowledge*, den gemensamma kunskapen, har blivit emellertid alltför perifer inom den kognitiva vetenskapen liksom inom lingvistiken. Saad och Maher (1996) inser, att begreppens gemensamma representation kan delas i flera områden, av vilka de tre viktigaste för designeryrket är den *systembundna* (bunden till fackliga system), den *semantiska* och den *teknisk-funktionella* representationen. Vi lingvister är mest intresserade av den semantiska representationen för begrepp, men vi måste dock minnas, att det finns andra representationer, som står i växelverkan med den traditionellt sett semantiska. I det specialiserade planeringspråket kan till exempel ett sådant vardagligt ord som *vägg* ha en komplicerad för en yrkesgrupp gemensam semantisk representation (vid sidan av den grafiska representationen).

*-- a wall might be represented in a [computer] system by double lines. -- In order to extend the communication beyond the visual representation of the*

*design, the content includes an explicit representation of design semantics. Design semantics represent the abstraction of design information which describe the purpose and performance of the design objects that go beyond the geometry to include such considerations as function, behaviour, versions, general design knowledge etc.* (Saad & Maher 1996)

Den semantiska representationen har alltså fått sina ingredienser av de yrkesverksamma personernas abstraheringsprocesser. Planerarna (arkitekterna, formgivarna) abstraherar främst sina intentioner som hänger samman med objektets funktion och konstruktion. För att handla med hjälp av den gemensamma kunskapen skapar man inom branschen ett gemensamt språk. Den potentiella gemensamma intentionen att skapa språk kan ta mycket svårförklarliga former såsom jag har konstaterat vid iaktagandet av formgivare. Även om man kunde förklara det epistemiska och det sociala i språkbrukssituationer på ett tillfredsställande sätt skulle den kognitiva situationen inte vara helt konstruerad ännu. Det återstår att granska situationen också som en individs tillstånd.[9]

*Vasa universitet*
*FIN-65101 VASA*

---

[1] Jfr Janik 1996: 22B23; von Eckardt 1995: 197B231.

[2] Se Jackendoff 1997: 540B541.

[3] Se Cruse 1988: 73.

[4] Se Karihalme 1996: 61B69 och 1998.

[5] Vihma 1995: 37B41.

[6] Se Cooper 1997: 65B86.

---

[7] Jackendoff 1997: 539.

[8] Se Dennett: 140B145.

[9] Jfr Cooper 1997: 76; von Eckardt 1995: 6B7.

## Litteratur

Britton, B. K. & Pellegrini, A. D. (eds.) 1990. *Narrative Thought and Narrative Language*. Hillsdale: Erlbaum.

Cooper, R. 1997. The Role of Situations in Generalized Quantifiers. I: S. Lappin (ed.) *Contemporary Semantic Theory*. Bodmin: Blackwell Publishers. 65B86.

Cruse, D.A. 1988. Word Meaning and Encyclopedic Meaning. I: W. Hüllen & R. Schulze (eds.) *Understanding the Lexicon*. Tübingen: Max Niemeyer Verlag. 73B83.

Dennett, D. C. 1997. *Miten mieli toimii* [Kinds of Minds]. Juva: WSOY

Edwards, D. 1997. *Discourse and Cognition*. Trowbridge: Sage.

Jackendoff, R. 1997. Semantics and Cognition. I: S. Lappin (ed.) *Contemporary Semantic Theory*. Bodmin: Blackwell Publishers. 539B559.

Janik, A. 1996. *Kunskapsbegreppet i praktisk filosofi*. Stockholm/Stehag: Brutus Östlings Bokförlag Symposion.

Karihalme, O. 1996. *Muotoilun teoriasanaston termistyminen*. Acta Wasaensia 51. Vasa: Vasa universitet.

- 1998. *Die Terminologisierung des Theoriewortschatzes des Design*. Opublicerad.

Keinonen, T. 1992. Käyttäjäliittymien muotoilu. Muoto 3. 60B63.

Kemmis, S. 1985. Action research and the politics of reflection. I: D. Baod & R. Keogh & D. Walker (eds.). *Reflection: Turning experience into learning*. London: Kegan Paul.

Kirsch, D. & Maglio, P. 1994. On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science* 18. 513B549.

Saad, M. & Maher, M.L. 1996. Shared understandingin computer-supported collaborative design. *Computer-Aided Design* 28. 180B192.

von Eckardt, B. 1995. *What is Cognitive Science?* USA: MIT Press.

Vihma, S. *Products as presentations. A semiotic and aesthetic study of design products*. Helsinki: University of Art and Design Helsinki.

# On grammaticalization of body-part nouns: *käsi* 'hand' in Finnish and Estonian

*Krista Ojutkangas*

### Abstract

In this paper I sketch out the grammaticalization of the inner local case forms of the word *käsi* 'hand' in two Balto-Finnic languages, Estonian and Finnish. The word still functions as a proper noun, but can also be an adverb and a postposition, where the main grammatical function is to express possession and state. There are two factors that adjust the possessive use of these forms: first, the choice of the predicate verb of the sentence; and second, the human qualities of the possessor. Both factors can be seen as remnants of the original meaning of 'human forelimb'. The same forms expressing state change the viewpoint from the possessive meaning.

## 1. Introduction

In this paper I examine the grammatical uses of the word *käsi* 'hand' in the Estonian and Finnish languages, where, in addition to the use as a proper noun, it functions as a postposition and an adverb, having a possessive or a spatial meaning. This presentation of the meanings and functions of the word *käsi* is by no means exhaustive, but I have restricted my study to the grammaticalized inner local case forms of the word.

There are two central factors in the development of the grammatical function of the word. First, an expression of location—holding something in one's hand—is a metaphorical source for an expression of possession. Second, the word's clear meaning as an expressly human body-part gives a strong anthropocentricity to the meanings of the word *käsi*: even the grammatical use of the word is often restricted to contexts with human participants. Especially in Finnish, the physical strength and skill of the human hand implies a balance of power in the possessive meaning of the word. Although locality is the central source of grammaticalization, the unique physical nature of the human hand has certainly affected the development of the word's meaning and functions. The effect of anthropocentricity on the grammaticalized meaning of the word shows that this functional development has not yet taken the word very far from its concrete source.

Estonian will be the starting point in this paper given the fact that *käsi* is clearly more grammaticalized in Estonian than it is in Finnish, a claim which will be returned to later. From the Finnish language I present some more or less idiomatic parallels of meaning. I wish to present the grammaticalization process of the word *käsi* as lexical extension. My view is that the mechanisms of grammaticalization do not differ from the mechanisms of language change in general: it is based on analogy, re-analysis, and extension. (See e.g. Harris and Campbell 1995: 92, 101—115.)

## 1. Possession

### 1.1. The form and the syntactic role of the word

As adverbs and postpositions, the central grammatical function of the inner local case forms of the word *käsi* is to express possession.[1] This kind of development is widely documented from languages all over the world (see e.g. Heine & al. 1991: 34, 53). An expression of the location of an object in the hand has become an expression of the holder's possession of the object, where the two concepts, location and possession, have been metaphorically connected in the mind of the speaker. The possessed item is kept in the hand, and the forms of the word 'hand' expressing possession are thus inner local case forms in both Estonian and Finnish.

In Estonian the use of the inessive case form *käes* ('in the hand'), expressing stylistically neutral possession, is not very common, and example (1) is one of the few in my database. In (2) *käes* expresses the power or ability of aunt Olga. Thus, the inessive case being rare, there seems to be a tendency in Estonian to express a change of the possession with the elative *käest* ('from the hand') and illative *kätte* ('into the hand') forms (see examples 7—10 below).

| | | | | |
|---|---|---|---|---|
| (1) Est. | *Siis* | *on* | *pilet* | ***käe-s.*** |
| | then | is | ticket | hand-sg.ine |
| | 'Then she has the ticket [in hand].' (Sclomb Ilu) | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| (2) Est. | ***Tädi Olga*** | | ***käe-s*** | *on* | *see* | *naljaasi* |
| | aunt Olga.sg.gen | | hand-sg.ine | is | it | play-thing |
| | *ära* | *korraldada.*(Sclomb Ilu) | | | | |
| | away | arrange | | | | |
| | 'For [in the hand of] aunt Olga, such a thing is very easy to arrange.' | | | | | |

In Estonian, the forms of *käsi* are singular in number, in Finnish, forms are typically plural. In Finnish, the forms *käsissä, käsistä, käsiin* do not purely express possession, but rather occupation, or control. This feature in the meaning follows of course from anthropocentricity: skill and power are in human hands. In Finnish, the singular form can have a kind of transitional meaning in the process of the grammaticalization of word, as in example (3). The plural *käsiinsä* in (4) is clearly more abstract, and it expresses acquisition, getting (the journal) into one's possession.

| | | | | | | |
|---|---|---|---|---|---|---|
| (3) Fin. | *Sitä paitsi* | *minusta* | *on* | *kohtalokasta,* | *että* | *mies* |
| | besides | I | be | fatal | that | man |
| | *saa nuoruude-ssaan* | | *kirjan* | ***käte-ensä.*** | (LA Plays 9921/503) | |
| | get youth | | book | hand.sg.ill.px | | |
| | 'Besides I think it is fatal that in his youth a man gets a book into his hand.' | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| (4) Fin. | *Miten* | *on* | *mahdollista,* | *että* | *marsalkka* | *on* |
| | how | be | possible | that | marshal | be |
| | *saanut* | ***käsiinsä*** | *tämän* | *lehden - -* | (LA Plays 9921/768) | |

get hand.pl.ill.px this journal
'How is it possible that the marshal has got this journal into his hands- -'

In the possessive meaning the syntactic role of the word is flexible in both Estonian and Finnish: it can function as an adverb (1, 3—4) as well as a postposition (2). This syntactic flexibility is a typical feature of the Balto-Finnic system of spatial expressions. The majority of the adpositions in Estonian and Finnish are in fact close to relator nouns and they still function as proper nouns as well; the emergence of new adpositions from nouns is a productive process in these languages (see e.g. Alhoniemi 1988, 28; DeLancey 1997).

## 1.2. Grammaticalization as lexical extension

### 1.2.1. The predicate verb

The Estonian occurences in my database can easily be divided into two groups upon examining the predicate verb of the clause. In the first group the forms *käest* or *kätte* are used with more or less concrete verbs of acquisition, as in examples (5—6). As you can see from example (6), in Estonian the physical distance of the participants in the situation is not a restriction for the use of this postposition. The elative form *käest*, (lit.) 'from the hand', simply expresses the source of the action of acquisition, without any concrete connection to the original meaning 'human forelimb'.

(5) Est. *Lubab kohe homme hommikul kõik väljasõidupaberid*
promise directly tomorrow morning all emmigration papers
*oma* ***kätte*** ***korjata*** - - (Sclomb Ilu)
own hand.sg.ill gather
'He promises to gather all the emmigration papers [into his hand] right tomorrow morning.'

(6) Est. *Peeter* **tellis** *naaberkülast* **sepa** **käe-st**
P. order neighbour-village smith.sg.gen hand-sg.ela
*endale* *kirve.* (Rätsep 197: 162)
himself axe
'Peter ordered himself an axe from [the hand of] the smith from the neighbour village.'

The word *käsi* in a possessive meaning is most concrete in contexts where we talk about tangible objects that can really be held in the hand. In Estonian, this is especially clear with predicate verbs that mean throwing, as in example (7). Clauses like these can be ambiguous: the illative form *kätte* 'into the hand' can in principle be understood concretely, but it also expresses that the action of throwing was successful. In other words, it has a function that is typical for adverbs defining verbs: it expresses perfectivity, as you can see from the modification in (7'). Verbs of throwing are a subgroup of verbs of acquisition.

(7) Est. *Laps* **viskas** *palli* **ema** **kätte**.(Rätsep 1978: 173)
child throw ball mother.sg.gen hand.sg.ill
'The child threw the ball to [the hand of] the mother [who caught it]'

(7') *Laps viskas palli* **emale**.
mother.sg.all
'The child threw the ball to the mother [and we don't know if she caught it]'

Like acquisition verbs, motion verbs are also possible predicates in the first group of occurences, as in the following examples (8—9). We can see these and examples (5—7) above as the starting point of the grammaticalization of the word *käsi*: we are talking about receiving, giving, or the movement of concrete, tangible objects.

(8) Est. *Kuhu nad kaovad* *ja* **kelle** **kätte** **sattuvad** - -
where they disappear and whose hand.sg.ill fall
'Where do they disappear and to whom [into whose hand] do they fall' (Sclomb Ilu)

(9) Est. *Vahepeal* *on* *aga* *võim* *täielikult*
between is but power wholly
***koristaja*** ***kätte*** ***üle-läinud***. (Sclomb Ilu)
cleaner.sg.gen hand.sg.ill over-go
'But in the meantime the power has entirely passed to [the hand of] the cleaner.'

In Finnish as well, the inner local case forms of the word *käsi* can be used as postpositions or adverbs with an acquisition verb; see examples (3—4) above. Motion verbs are also possible predicates in Finnish, as in the following examples (10—11).

(10) Fin. *Paljon vaikeampaa se oli silloin, kun Jeesus **joutui** Getsemanessa*
Jesus fall.into G.
***vihollistensa*** ***käsiin***. (LA Devotional literature 9781/321)
enemy.pl.gen.px hand.pl.ill
'It was much more difficult when Jesus got into the hands of his enemies in Getsemane.'

(11) Fin. *Aika **valuu käsistä** niin kuin hiekka tiimalasissa.*
time flow hand.pl.ela
'The time passes [from the hands] like sand in the hourglass.' (LA Magazine reports 9611/599)

In Estonian, in the second group where the possessive *käes*, *käest*, *kätte* occur, the predicate is a speech act or a cognitive verb. The following examples illustrate this use, which is met only in Estonian—it's not linguistically possible in Finnish at all. This represents a more abstract, i.e. more grammaticalized, side in the use of the word *käsi* in the possessive meaning: typically the question is of acquiring or possessing information (12—13), knowledge or skill (14).

(12) Est. *Tagumised* ***pärivad*** ***esimeste*** ***käe-st*** - - (Sclomb Ilu)
rearmost ask first.pl.gen hand-sg.ela
'Those standing in back ask from [the hand of] the others in front --'

(13) Est. *Lion* -- ***korraldab*** ***nüüd*** *ise* -- *ilma*
L. settle now self without
***ema*** ***käe-st*** *abi* ***küsimata***
mother.sg.gen hand-sg.ela help ask
*ka* *tädi* *Olga* *väljasõidu*. (Sclomb Ilu)
also aunt O. emmigration
'Lion now arranges also aunt Olga's emigration on his own, without asking for help from [the hand of] mother.'

(14) Est. *Peeter* ***õppis*** ***meistri*** ***käe-st*** *ameti*. (Rätsep 1979: 81)
P. learn master.sg.gen hand-sg.ela profession
'Peter learned a profession from [the hand of] the master.'

### 1.2.2. The requirement for human or animate possessor

When possession is expressed with the word *käsi*, the possessor must be human, since only human beings have hands. As an exception, Estonian sometimes allows the possessor to be an animal, as in example (15). The use of animal possessors is constrained by the condition that the animals must possess a certain level of dexterity, like dogs or cats. It is not possible, for example, 'to give anything to the hand of a cow'. Thus, even here, there is something left of anthropocentricity, the skillfullness of the human forelimbs. When the meaning of the body-part noun moves away from its origin, inanimate possessors become possible, as in example (16), but this is still rare. In Finnish, the demand for a human possessor is stronger than in Estonian, but in addition to an individual human being, the possessor can be a community or group of people, as in (17).

(15) Est. *Hannes* *pillab* *raasukesed* *taldrikult*
H. drop crumbs plate
*laua* *alla* ***koera*** ***kätte*** (Rätsep 1978: 173)
table under dog.sg.gen hand.sg.ill
'Hannes dropped the crumbs to [into the hand of] the dog under the table.'

(16) Est. *-- astub ta katkise elektripirni peale, mille* ***laine*** *on*
step s/he broken electric-lamp on which wave is
*kusagilt* ***kätte*** *saanud ja kaasa toonud.* (Sclomb Ilu)
somewhere hand.sg.ill get and with bring
'She steps onto a broken electric lamp that a wave has got [into its hand] and taken with from somewhere.'

(17) Fin. *Tiedot löytyvät presidentti Urho Kekkosen muistiinpanoista, jotka*
*Norjan radio NRK on saanut* ***käsiinsä***. (LA Newspaper articles 9522/3)
Radio of N. be get hand.pl.ill.px
'The information can be found in president Urho Kekkonen's notes that the Radio of Norway NRK has got [into its hands].'

As a summary, we can sketch the development of the possessive meaning of the forms *käes, käest, kätte* in Estonian as a lexical extension (see Table 1). We can see how the possible contexts have grown from the original concrete holding of an item in one's hand: The group of possible predicate verbs has extended from concrete acquisition and motion verbs to speech act and cognitive verbs. From this it follows that abstract entities like knowledge and skill have become acceptable as possessees. As the connection with the original meaning of the human body-part fades away, the use of the adverb-postposition extends to cases with non-human and inanimate possessors.

Table 1. The possessive meaning of the Estonian adverb-postposition *käes, käest, kätte*: a summary of the lexical extension.

| **The part of the structure under the condition** | **The nature of the lexeme before extension** | | **The nature of the lexeme after extension** |
|---|---|---|---|
| The predicate verbs | acquisition and motion verbs | | speech act and cognitive verbs |
| The possessees | (tangible) objects | | skill or knowledges |
| The possessors | human beings | animals | inanimates |

## 2. State: a change of viewpoint

In Estonian the inner local case forms of *käsi* are also used as expressions of state, as in the following examples (18—20). What is the relationship between this and the possessive meaning?

(18) Est. *Miina piinles* **valu** **käe-s**. (Rätsep 1978: 94)
M. suffer pain.sg.gen hand-sg.ine
'Miina suffered from [in the hand of] the pain.'

(19) Est. *Eriti haaratud on ta aga praegu oma*
especially impress is she but now own
***kuumuse*** ***käe-s*** *õhetama löönud näonahast* - - (Sclomb Ilu)
heat.sg.gen hand-sg.ine glow hit face-skin
'But she is especially impressed by her face that has started to glow in [the hand of] the heat.'

(20) Est. *Liiv levitab kuumade, kaua* ***päikese*** ***käe-s*** *seisnud*
sand spread hot long sun.sg.gen hand-sg.gen stand
*pliiatsite ja kuiva kollaka paberi lõhna*. (Sclomb Ilu)
pencil anddry yellow paper smell
'The sand spreads the smell of warm pencils, that have long laid in [the hand of] the sun, and the smell of dry yellow paper.'

This relationship can be seen simply as a change of viewpoint. The point is that being possessed is a state. We can see being in a state as being in the possession of someone or something. Again, we find that the power of the possessor expressed by the word *käsi* is an essential feature of the meaning. The state is typically a negative feeling or an illness, as in example (18), or it can be a more general condition like weather, light or temperature, as in examples (19—20). All of these unavoidably have the power over the persons or entities concerned. Metaphorical personification can be seen as the starting point for this meaning, but note that in the language at the moment these expressions of state are neutral in style and need not be understood figuratively.

When used to express state, the form of the word *käsi* is the same as the one expressing possession, but here it is clearly a postposition that requires a genitive adjunct. The state is of course inanimate. These two factors - the status of a postposition and lack of the requirement of animateness - may give us a reason to consider the expressions of state as more grammaticalized than the expressions of possession.

## 4. Conclusion

In the following, I will summarize the evidence for the claim, shown in the analysis of the examples listed above, that the inner local case forms of the word *käsi* are more grammaticalized in Estonian than in Finnish.

As expressions of possession, the inner local case forms of the word *käsi* can be used as a postposition or an adverb with acquisition and motion verbs in both languages. However, speech act and cognitive verbs are possible as the predicate only in the Estonian language. From this it follows that, in Estonian, the possessed entity can be knowledge or skill, which is not possible in Finnish. In the prototypical case, the possessor that is expressed by the word *käsi* is human. In Estonian the possessor can in some cases be an animal or even inanimate, but this is still very rare. In Finnish, the requirement for a human possessor is stronger. Thus in Finnish the word is clearly closer to its concrete original meaning of 'human forelimb' than in Estonian. The contexts of its occurences in Finnish are perhaps closer to idiomatic than productive.

As expressions of state, the inner local case forms of the word *käsi* are possible in practice only in Estonian. In Finnish, this type is very marginal, although it does exist in a few idioms. From the meaning of state it follows that the possibility of inanimate participants increases; state is always inanimate and the entities in the state can be either animate or inanimate. In the expressions of state the syntactic solidity of the structure increases, since the forms become postpositions that require a genitive adjunct. In this context, the inner local case forms of the word *käsi* have become a grammatically obligatory part of the structure.

Finally, as a general point, we note the frequency of the forms examined in the two Balto-Finnic languages. In the part of my database that derives from the Sclomb corpus (i.e. from two novels), there are 60 Estonian grammatical occurences unambiguously expressing either possession or state of the inner local case forms of the word *käsi*. In the Finnish counterparts of the texts, there are considerably fewer (only 4).

*Department of Finnish*
*Fennicum, Henrikinkatu 3*
*FIN-20014 University of Turku*
*Finland*
*email: krioju@utu.fi*

## Notes

1. Note that in the prototypical possessive structure of Estonian and Finnish, the possessor is expressed with an outer locative case, the adessive (lit. 'on, at'), and the possessor and the possessee are connected with the copular verb 'to be':

| | | | | | |
|---|---|---|---|---|---|
| (1) Est. | Minu-l | on | uus | raamat | |
| Fin. | Minu-lla | on | uusi | kirja | |
| | I-sg.ade | be.pr.sg.3 | new | book | 'I have a new book.' |

Abbreviations used in the glosses

| | | | |
|---|---|---|---|
| ade | adessive case | ine | inessive case |
| all | allative case | pl | plural |
| ela | elative case | px | possessive suffix |
| gen | genitive case | sg | singular |
| ill | illative case | | |

## References

Alhoniemi, Alho 1988. Postpositiorakenteiden synkroniaa ja diakroniaa. [On synchrony and diachrony of the postpositional structures.] Turku: Sananjalka 30.27—44.

DeLancey, Scott 1997. Grammaticalization and the gradience of categories: relator nouns and postpositions in Tibetan and Burmese. Essays on language function and language type. Dedicated to T. Givón, ed. by Bybee, Joan, John Haiman & Sandra A. Thompson, 51—69. Amsterdam: John Benjamins Publishing Company.

Harris, Alice C., and Lyle Campbell 1995. Historical syntax in cross-linguistic perspective. Cambridge: Cambridge University Press.

Heine, Bernd, Ulrike Claudi, and Friederike Hünnemeyer 1991. Grammaticalization. A Conceptual Framework. Chigago and London: The University of Chigago Press.

LA = The Syntax Archives. Department of Finnish and General Linguistics. University of Turku.

Rätsep, Huno 1978. Eesti keele lihtlausete tüübid. [The basic sentence types of Estonian.] Eesti NSV Teaduste Akadeemia. Emakeele Seltsi toimetised nr. 12. Tallinn: Valgus.

Sclomb = Studia comparativa linguarum orbis Maris Baltici. Corpus for the contrastive study of the Baltic Sea region languages. Department of Finnish and General Linguistics. University of Turku.

------ Ilu = Luik, Viivi 1991. Ajaloo ilu. Tallinn: Eesti Raamat. Novel, part of the Sclomb corpus.

# Some pragmatic properties of French cleft constructions

*Eija Suomela-Salmi*

## Abstract

This paper discusses some pragmatic properties of cleft construction in written French discourse. Consequently the paper focuses on c'est-clefts coresponding to it-clefts in English and on so called pseudo-clefts coresponding to wh-clefts in English only very briefly touching upon the so-called presentational clefts typical of spoken French discourse. The aim of the paper is to examine to what extent the relevance theoretical account offered by Jucker (1997) for English is applicable for equivalent constructions in French. Finally some tentative answers considering the distribution of these constructions in different types of discourse are offered.

## Introduction

The aim of this paper is to examine whether the explanation Jucker (1997) offers for cleft constructions and the so called pseudo-cleft constructions in English does apply to French as well. In French these constructions are described traditionally as being emphatic – 'construction de mise en relief, de mise en valeur' - (Grevisse & Goosse 1993:695), serving the expressive needs of the speaker (Blinkenberg 1928:33), belonging to the affective syntax of the speaker (Bally 1909:250) and consequently, bearing specific stylistic values. Clefts have also been discussed abundantly within the framework of generative grammar (for French cf. Dubois & Dubois-Charlier 1970, Le Galliot 1975, Gross 1977) and textlinguistics (cf. for French e.g. Combettes 1988, Galmiche 1992) which offers largely contradictory views dealing with the information structure of this bipartite construction.

My approach, however, is pragmatic i.e. in this paper I concentrate on the relationship between the syntactic constructions in question and the cognitive and communicative constraints affecting their use.

The focus will be on cleft-sentences (phrases clivées) such as

1. C'est Charlotte qui aime le chocolat.

and the so called pseudo-clefts (phrases pseudo-clivées)

2. Ce que Charlotte aime c'est le chocolat.

even if the term 'pseudo-clivée' has been stongly criticised in French linguistic literature. Valli (1981:195), for example, argues that there is no transformation in French called 'pseudo-clivage' and consequently, pseudo-clefts are in fact basic constructions in French which cannot be derived from a non-clefted sentence. He states that a better description of the structure in question would be the following approximative schema: 'A c'est B' = 'A it is B'. Valli among others considers that cleft constructions in general are not governed solely by morphosyntactic constraints but they should rather be seen as a discursive configuration.

Below I will only very briefly discuss what Lambrecht (1987,1988) calls presentational cleft constructions especially frequent in spoken French of the type: [y + a NP qui VP] rouhgly corresponding to English 'there is NP who'...

3. Y'a Jean qu'a téléphoné.

or

[clitic + a NP qui VP]

4. J'ai les yeux qui me font mal.

## Presentational cleft constructions

According to Lambrecht (1987,1988) in spoken French presentational cleft constructions are motivated by the tendency to remove full lexical NP's with focus function away from the preverbal subject position and to introduce discourse referents not yet available (either brand-new or unused in the sense of Prince 1981) in preferred post-verbal position. There seems to be an apparent similarity with these presentational constructions and the so called Discourse processing function of Left-Dislocations (one of the three functions assigned to Left Dislocations by Prince 1998). According to Prince (1998:124) 'a Simplifying Left-Dislocation serves to simplify the processing of Discourse-new entities by removing them from a syntactic position disfavoured for Discourse-new entities and creating a separate processing unit for them. Once that unit is processed and they have become Discourse-old, they may comfortably occur in the positions within the clause as pronouns'.

It should be pointed out that the [(il) y a NP qui VP] cleft construction can only be used existentially, the corresponding deictic presentational construction being [voilà NP] as in the following example:

5. Voilà Jean qui arrive.

The main pragmatic difference between the two is that the deictic presentational constructions (examples 4 & 5) point to a referent in the text-external world, which comprises speech participants and speech setting, whereas the existential 'il y a' construction (example 3) introduces a referent into the text-internal world consisting of the linguistic expressions and their meanings (Lambrecht 1994: 36-37, 179). However, whether deictic or existential, all presentational constructions have in common the basic feature that they assert the presence of the referent in the text (external or internal) world. This newly introduced referent is merely located in the discourse or made available for predication in subsequent clauses by being raised into the adressee's consciousness (idem: 180-181).

## Some properties of clefts, pseudo-clefts and their non-cleft equivalents.

Cleft constructions and their corresponding clefted variants (cf. Lambrecht's allosentences, i.e. semantically equivalent but formally and pragmatically divergent surface manifestations of given propositions, 1994:35) are truth-conditionnaly identical. Thus if

a) Charlotte aime le jupe rouge

b) C'est le jupe rouge que Charlotte aime

c) Ce que Charlotte aime c'est le jupe rouge

are taken to refer to the same Charlotte and to the same red skirt, then all of these sentences must be either true or false at any given moment.

The difference between a) and b)-c) is that it is the red skirt that is presented as particularly important in the sentence either in the sense that it is unpredictable information or that it exclusive in some sense.

When analysing cleft constructions we should, according to Jucker (1997) and Lambrecht (1994) ask ourselves at least the following questions:

a) why does the speaker/writer use a cleft sentence rather than its non-cleft equivalent?

b) what is the difference between the different types of cleft constructions?

Jucker (1997:188) adds one more question:

c) how does this difference account for the different distributions of these constructions?

As for the first question, Jucker offers a relevance theoretical account according to which 'cleft sentences should be seen as providing procedural rather than conceptual meaning. They should be seen as processing instructions to the adressee making explicit which elements of the sentence should be processed as background and which elements as foreground information' (ibid).

As for the second question: What is the difference between these cleft constructions? Jucker (1997:192) answers that it-clefts are foregrounding devices wheras wh-clefts are backgrounding devices irrespective of the information status of the highlighted element.

Lambrecht (1987:224-225) offers a differently formulated explanation of clefts and their corresponding allosentences. He calls French 'c'est-clefts' - corresponding to it-clefts in English - identificational or focus marking constructions since a constituent whose referent plays the pragmatic role of focus is allowed to appear in the post-verbal position.This is done by creating an auxiliary clause containing the copula 'être' in whose free focus position the constituent can appear. In the 'c'est-cleft construction', it is the predicate that is pragmatically recoverable and the clefted NP that adds the new, unpredictable or otherwise non-retrievable piece of information. Thus the 'c'est-cleft' is, in a sense, reverse to the pragmatically unmarked sequence in which a topical element is followed by a predicate containing new or unrecoverable information about this element.

In this sense making the choice between the cleft and its non-clefted allosentence would depend on which part of the sentence is presented as presupposed and which part as not yet accessible information to the adressee. This is combined with the pragmatically motivated syntactic constraint of speakers favouring a position when introducing not yet active discourse referents - object position - whereas the subject is strongly disfavoured as a site for Discourse-new referents (brand-new or unused).

The main difference between the accounts of Jucker and that of Lambrecht of the same phenomenon, c'est-clefts, lies in the emphasis given to the cognitive status of the utterance interpretation which emphasizes the interpretative interplay inherent in any message vs. the sender orientated view and the part played by syntactic structures which are discourse linguistically motivated i.e. 'the relationship between the formal structure of sentences and the communicative situations in which sentences are used to convey propositional information' (Lambrecht 1994:334). The first one is process orientated, the latter to a large extent product orientated, and yet pragmatic both of them.

## A brief relevance theoretical survey

Sperber and Wilson (1986/1995) argue that the way hearers interpret utterances is restricted by a particular feature in their cognitive make-up: humans look for relevant information which yields the maximum possible changes in their cognitive environment for the minimum possible processing effort. Hearers look for an interpretation which is optimally relevant, ie. an interpretation that yields enough effects to be worth processing without putting the hearers to an unjustifiable effort in deriving them and which is compatible with the speaker's preferences. As soon as the hearer finds such an interpretation that the speaker might reasonably have been intending to communicate, the hearer assumes this is the right one and looks no further.

This means that all elements linguistic or not must either add to the effect side ('yield adequate effects for the minimum justifiable effort' (Sperber &Wilson 1993:7) that is to say make the interpretation worthwhile) or they must help to reduce the effort needed for the interpretation process. One way of reducing the necessary processing effort is to give the adressee specific processing instructions.

Interpreting an utterance consists of deriving its explicitly communicated content and its implicitly communicated import. The explicatures are according to Sperber and Wilson (1993:14) explicitly communicated assumptions. 'Every explicature is recovered by a combination of decoding and inference, and the greater the element of decoding, the more explicit will it be' (ibid). Implicatures, on the other hand, are arrived at entirely inferentially. Linguistically encoded information may contribute to the interpretation process either by forming part of a decoded conceptual representation or by providing information on the manipulation of conceptual representations eg. by offering guidance for inferencing. Most words encode concepts which form part of larger conceptual representations. Most words which encode concepts contribute to the assertions made by the utterances in which they occur by affecting their truth conditions.

Some words encode procedural information i.e. information on how to manipulate, how to take a conceptual representation. Procedural information is best understood as a way of constraining or guiding the inferential phase of communication. In the relevance theoretical perspective, utterance interpretation is largely an inferential process which consists of forming and evaluating hypothesis on what the speaker intends to communicate. Procedural information restricts the number of hypothesis the hearer needs to consider in order to arrive at an optimally relevant interpretation, thus facilitating his task. (Rouchota 1998: 110-111, Sperber &Wilson 1993:10-11).

Jucker (1997:189) argues that the clefts give the adressee more specific processing instructions than their non-cleft equivalents. In a non-cleft sentence, the hearer has to establish the focus of the sentence as part of the interpretation process, in particular what is foreground information and what is background information which is only indirectly relevant. Thus cleft-sentences should be seen as a means of focus assignment as also pointed out by Lambrecht (1987:1994).

## Analysis

### 1. C'est-clefts

According to Jucker it-clefts - corresponding to French 'c'est-clefts'- signal foregrounded information to the reader irrespective of the information status of the clefted entity.

1. A l'approche du village, Marina ôtait vite vite ses bijoux fantaisie et les fourrait dans son sac avec le badge qu'elle affectionnait et qu'elle épinglait à son chemisier, libera ma non troppo, *c'était sa devise qui* était gravée dessus,

proclamait-elle. (Daniel Maggetti, Les grâces, Le passe-muraille, Journal littéraire no 37, juillet 1998:2).

Clearly, the referent designanted in the clefted sentence ( the slogan *libera ma non troppo* on her badge) is accessible to the reader via the immediate preceding context, but this is not the crucial piece of information. What the reader is expected to interpret as the foregrounded / most important information is that Marina considered this phrase as her motto.

2. *C'était alors*, sur un guéridon, *que* je remarquai un livre apparemment oublié: La Mort en Eté de Mishima. (Asa Lanova:Lettre d'Alexandrie,Le passe-muraille, Journal littéraire no 37, juillet 1998:6)

The same argument applies for the second example. The temporal expression clefted *alors* - it was then that I noticed a book apparently forgotten - is textworld internally easy to access, the point here is to emphasize the unexpected or exclusive temporal moment of the event.

3. J'aime une femme, disons que *c'est plutôt elle qui* est tombée amoureuse de moi. Et moi je la regarde tomber. Elle commence à tomber. Elle met un tel temps à tomber qu'elle s'efface avant de toucher terre. Et *c'est moi qui* m'effondre, brisé en mille morceaux, sur le parvis de la place…(Ann-Lou Steininger: Avec la musique lentement, Le passe-muraille, Journal littéraire no 37, juillet 1998:9).

In this example the unexpected  information is clearly foregrounded and furthermore emphasized by the adverbial *plutôt* even though the referent of *elle (she)* is easily accessible. In the second cleft *moi* (me) is not as easily accessible since a large textual chunk separates it from it's previous occurrence and the main discourse topic has for some time been *elle (she).* In this case the unexpected, even if not totally unused information, is highlighted via the cleft in order to facilitate the processing of the contrast indicated in the relation *elle-moi* (she-me) which caracterizes the entire short story.

4. En revanche, quand il s'agit non d'œuvres singuliers, mais de rapports de stage, de reportages sportifs télévisés, de cours d'université, etc. il n'y a pas d'affiliation à l'égard d'œuvres consacrées: *ce sont des routines, des comportements stéréotypés et anonymes qui se sont stabilisés peu à peu mais* sont sujets à variation continuelle …Les genres de discours ne sont pas des sortes de moules dans lesquels le locuteur choisirait de couler son énoncé. En fait, *ce sont des activités sociales qui* sont en tant que telles soumises à un critère de réussite. (Maingueneau: Analyser les textes de communication 1998: 51. Paris. Hachette.)

In example 4 taken from an academic textbook the relative clause of the first cleft - *qui se sont stabilisés peu à peu* - expresses information that is presented as presupposed since in the case of the developement of discourse genres we would expect a certain stabilisation process to take place whereas *routines and stereotyped anonymous behaviour* is the crucial new and highlighted information. In the second cleft, the clefted entity – *des activités sociales* – reclassifies the previously mentioned referent - *les genres de discours* – discourse genres, thus contributing to the new and contrastive interpretation aimed at . The discourse marker *en fait* at the beginning of the sentence further emphasizes this contrastive interpretation.

5. Cotexte. *Ce sont les séquences verbales qui* se trouvent placées avant ou après l'unité à interpréter. (Maingueneau 1998:12)

5'. *Le séquences verbales qui se trouvent placées avant ou après l'unité à interpréter sont le contexte.

5'' Cotexte. Les séquences verbales qui se trouvent placées avant ou après l'unité à interpéter.

Why does 5 ' sound at least strange if not unacceptable? Probably, because of the tendency discussed earlier, to avoid brand-new referents in subject positions, even if the canonical SVO order would allow it. There is no way in which the full lexical NP

– 'les séquences verbales' - could have become salient in the text world since the sentence following the term 'cotexte' is a definition of it. The principle or pragmatic maxim proposed by Lambrecht (1987:254) seems to apply: 'Don't introduce a referent and talk about it at the same time'.

In 5'' this maxim is not violated. The mere juxtaposition of these two utterances is enough to separate the new discourse referent and what is said about it. Furthermore as Valli (1981:207) points out, there is a certain ressemblance between specifying clefts of the type 'A c'est B' and other specifying discourse structures and sometimes there is nothing to introduce this specifying structure. In such cases there is heavy reliance on the addressee's capacity to process the information thus presented.

> 6. Cette pièce est un espace non-fumer et *cela fait déjà une heure qu'il* a envie d'en griller une et de siroter un bon whisky (Maingueneau 1998:12).

Example 6 is a modified version of the c'est-cleft with another intranstive verb 'faire' and hence the pronoun 'cela' as the pronoun 'ce' is only compatible with the copula 'être'. The semantic content of the verb 'faire' - make - has completely faded away in this construction. The information of the relative clause contains unused but through semantic contiguity accessible information (the antithetic relation of non-fumer et d'en griller une – to smoke one) The foregrounded and crucial information however in this case is that the purgatory has already lasted one hour and that no end to sufferings is in sight.

## 2. Pseudo-clefts

According to Jucker, the beginning of the relative clause in pseudo-clefts, which Jucker calls wh-clefts in line with Prince 1976, is presupposed and the rest of the construction presents a sort of specification of the presupposed content of the beginning. Pseudo-clefts present backgrounded information at the beginning of the relative clause and foregrounded information towards the end in the highlighted position. As Jucker (1997 :193) points out, this is the same distribution as in normal non-clefted sentences but in contrast to the non-cleft version the scope of the focus is not left open,but made explicit.

> *7. Ce dont* il s'agit ici, *ce n'est* pas de neutraliser le discours, d'en faire le signe d'autre chose et d'en traverser l'épaisseur pour rejoindre ce qui demeure silencieusement en deça de lui, *c'est* au contraire de le maintenir dans sa consistence, de le faire surgir dans la complexité qui lui est propre (Foucault: L'archélogie du savoir, 1969:65. Paris.Gallimard)

In this case the specification is presented as a series of negative affirmations thus narrowing down the range of possible interpretations and functioning as background information against which the information expressed with the c'est-cleft is presented as contrastive/highlighted and foregrounded information.

> *8. Ce qui* rend ce texte plus ou moins obscur à beaucoup d'étrangers, *c'est* la difficulté de l'attribuer un genre qui leur soit familier (Maingueneau 1998:50)

The background character of the beginning of thiS pseudo-cleft is further emphasized by the presence of the anaphoric or textually deictic 'ce texte'. The specification or the higlighted information is in this case unused but inferrable due the semantic association between obscurity and difficulty.

> 9. *Mais ce qui me frappa, surtout, ce fut* l'apparition du lac Karoun et, encore agitées par le khamsin, de ses eaux d'un gris houleux. (Asa Lanova: Lettre d'Alexandrie, Le passe muraille no 37, juillet 1998:6)

In example 9 we (the readers) already know that something had made an impression on 'me' and the last part of the pseudo-cleft *-ce fut l'apparition du lac Karoun -* specifies it making all other interpretations impossible. In this case the specification contains new information.

The text-referencing function of pseudo-clefts advocated by Jucker is illustrated by the following example:

> 10. Le parti communiste encore, comme un seul homme cette fois, qui ralliant dès 1935 l'essentiel des positions qui, à peine six mois plutôt, avaient suffi à exclure Doriot, en appelle maintenant à son tour à la grande alliance du peuple de France, tout entier mobilisé contre les voleurs et les parasites, tout entier rassemblé dans la communauté nationale resuscitée. Tant et tant de textes enfin, d'innombrables articles et discours où Vaillant-Couturier, mais d'autres aussi, trouveront les accents du chauvinisme le plus eculé, parfois même de la xenophobie la plus ignoble pour chanter la gloire de Jeanne-la-Paysanne, glorifier l'honneur et le parfum de nos terreurs, fustiger l'anti-France aussi et le cosmopolitanisme dissolvant des intellectuels bourgeois. *Ce qu'*on a oublié, autrement dit, *c'est que c'est* le pays tout entier de la droite à gauche, de la gauche à l'extrême gauche, de l'extrême gauche à l'extrême droite qui, cinq ans avant Pétain, communiait dans le même cri et rauque et déjà meurtier: 'La France aux français!'
> (Bernard-Henry Levy: L'idéologie française, 1981:32. Paris.Grasset)

The first part of the pseudo cleft – *ce qu'on a oublié* - summarizes and reformulates in fact the whole of the previous lengthy stretch of discourse by paraphrasing it and stating that what had been forgotten was the following – and the rest of the construction goes on to specify the reformulation. Jucker (1997:196) suggest that pseudo-clefts with text-referencing function and the processing help they offer are particularly suitable for spoken discourse. On the contrary, their presence in written discourse might be felt as patronizing because the readers have more time at their disposal to process the message. This claim,however, does not seem to be the valid for French written discourse.

The text-referencing function is not only limited to pseudo-clefts but is very frequent with c'est-clefts containing anaphoric expressions as in the following:

> 11... il faut aussi activer dans notre mémoire deux scripts: celui de l'émission de radio et celui de la drague. C'est le premier qui nous permet

> de faire un pont entre les deux premiers phrases… *C'est donc en activant ces deux scripts et tous les savoirs afférents qu'on* peut imaginer les ”quiproquos que cette situation peut provoquer” (Maingueneau 1998:28-29).

The first c'est-cleft is ’c'est le premier qui’ is assimilated via semantic contiguity as a subset of deux scripts whereas the second *c'est donc en activant ces deux scripts et les savoirs afférents qui* summarizes the whole of the previous stretch of text.

## Concluding remarks

The aim of this study has been twofold. Firstly, I wanted to find out whether the claims concerning the role of cleft constructions as providing guidance for the adressee advocated for English, would apply as well to French. Secondly, I was interested in finding an answer to Jucker's question how the difference between different cleft constructions accounts for the different distribution of these constructions.

The analysis revealed that cleft constructions in French discourse can be expalined by a relevance theoretical account. This is to say that they can be interpreted as carrying procedural meaning and thus acting as guidance for the readers as to which parts of the utterance are to be taken as foregrounded / most crucial information, and which as presupposed background information. The analysis revealed also that the information status of the clefted constituent does not play an equally crucial role as the foregrounding - backgrounding distinction, since the foregrounded parts often include easily retrievable information and vice versa, the backgrounded elements can contain unused, even brand-new information. It is of course healthy to bear in mind that the relevance theoretical account is only one of the many possible explanations to account for cleft constructions.

The answer to the second question is so far only tentative. First of all, the cleft constructions analysed in this study can appear both in written and in spoken discourse whereas the so called presentational clefts are clearly a feature of spoken French discourse. Jucker (1997) following Collins (1991) suggests that It-clefts corresponding to c'est-clefts, are especially a feature of persuasive discourse. So far, we have not found clear evidence to support this hypothesis in Frech discourse. It may be the case, especially if the concept of persuasion is understood in a very wide sense. These constructions seem to be also very frequent in academic textbooks and scientific texts. But can academic textbooks and scientific discourse be considered persuasive ? In such discourse types the function of pseudo-clefts again seems to be either ' A c'est B' i.e. to specify and thus to narrow down the possible range of interpretations concerning 'A' or the pseudo clefts have a text-referencing function summarising and thematizing a previous stretch of discourse. Is not specification rather a process of clarifying than convincing ? Firstly, since there are no monotype texts or discourses in real life, why should academic discourse not at least implicitly contain persuasive elements even if their dominant function were expository or informative. Secondly, in specifiying and thus narrowing down the possible range of interpretations certain choices have to be made : what to include and what to exclude. These choices again are made to convince the adressee of the relevance of the arguments and facts used.

Contrary to what Jucker claims, the use of pseudo-clefts as a text-referencing device does not in French written academic discourse seem to create a patronizing effect on the reader. This might be due to the relative verbosity and structural richness of French academic and sciëntific discourse which adds to the reader's processing load. Consequently this kind of processing information provided by pseudo-clefts is welcome. On the other hand, in academic text books a certain degree of partronizing is tolerated for the sake of clarity.

*University of Turku*
*French Department*

## References

Bally, Ch. 1909. Traité de stylistique française. Heidelberg : Carl Winter's Universitätsbuchhandlung.

Blinkenberg, A. 1928. L'ordre des mots en français moderne. Kobenhavn : Bianco Lunos Bogtrykkeri.

Collins, P.1991. Cleft and Pseudo-Cleft Constructions in English. London : Routledge.

Combettes, B. 1988. Pour une grammaire textuelle. La progression thématique. 2 nd édition. Bruxelles:De Boeck-Duculot.

Dubois, J. & Dubois-Charlier, F.1970. Eléments de linguistique française :syntaxe. Paris : Larousse.

Le Galliot, J. 1975. Description générative et transformationnelle de la langue française. Paris : Nathan.

Galmiche, M. 1992. Au carrefour des malentendus: le thème. in : L'information grammaticale 54, pp.3-10.

Grevisse, M & Goosse, A.1993. Le bon usage. Grammaire française. 13 e édition revue et refondue par André Goosse. Paris : Gembloux.

Gross, M. 1977. Grammaire transformationnelle du français. Syntaxe du nom. Paris: Larousse.

Jucker, A.H. 1997. The relevance of cleft constructions. in : Multilingua. Journal of Cross-Cultural and Interlingual Communication.Volume 16. 273, pp. 187-198.

Lambrecht, K. 1987.On the status of SVO sentences in French discourse. in : Tomlin, J.(ed.) Coherence and Grounding in Discourse, pp. 217-261. Amsterdam : John Benjamins.

Lambrecht, K. 1988. Presentational cleft constructions in spoken French. in Haiman, J. & Thompson, S.A. (eds.) Clause Combining in Grammar and Discourse, pp.135-180.

Lambrecht, K. 1994. Information structure and sentence form. Cambridge : Cambridge University Press.

Prince, E.F. 1978. A comparison of wh-clefts and it-clefts in discourse. in : Language 54, pp.883-906.

Prince, E.F. 1981. Toward a Typology of Given-New Information. in :Cole, P. (ed.): Radical Pragmatics, pp.233-255. New York : Academic Press.

Prince, E.F. 1998. On the Functions of Left-Dislocation in English Discourse. in : Kamio, A. (ed.) Directions in Functional Linguistics, pp.117-144.

Rouchota, V. 1998. Procedural Meaning and Parenthetical Discourse Markers. in: Jucker, A.H. & Ziv Y. (eds.) Discourse Markers, pp. 97-126. Amsterdam: John Benjamins.

Sperber, D. & Wilson, D. 1995. Relevance. Communication and Cognition. 2 nd edition. Oxford : Blackwell.

Valli,A. 1981. Note sur les constructions dites 'pseudo-clivées' en français. In : Recherches sur le français parlé no 3, pp. 195-211.

Wilson, D. & Sperber, D. 1993. Linguistic form and relevance. in. Lingua 90 (1/2), pp.1-26.

# The construal of the scene: Selected existential constructions

*Piotr Twardzisz*

**Abstract**

This paper addresses a topic which has a well-established tradition in linguistic investigations, namely that of existential constructions. The analysis is carried out from the standpoint of COGNITIVE GRAMMAR. The category of existential sentences is viewed here as a marginal instantiation among all processes. With preference for imperfective verbs and setting-like subjects, existentials depart from prototypical action-chain-like processes with two fully-fledged participants. The language material chosen comes from a few Germanic languages, specifically, from English, Swedish, Norwegian, German and Icelandic. The claim made is that existential constructions in the Germanic languages differ in the construal of the scene. Therefore the differences are semantic in character, rather than syntactic.*

It has been noted that the syntactic properties of existential sentences in the various Germanic languages differ considerably (see, for example, Falk 1993, Maling 1987, Platzack 1983). As for the syntax of existentials, researchers' attention has concentrated on the following three issues:

1. The kind of verb used, i.e. whether it is a transitive or intransitive one.
2. The status of the dummy pronoun, i.e. whether it is the grammatical subject of the sentence or not.
3. The indefiniteness effect in the logical subject NP.

Even the Germanic languages alone exhibit essential discrepancies as far as the kind of verb and subject of existential constructs is concerned. Keeping in mind the unavoidability of certain differences, even crucial ones, it is yet customary to seek common features shared by exemplary existentials in order to establish their characteristic parameters. For instance, according to Platzack (1983: 80): 'In the prototypical case, the existential sentence contains an intransitive verb, an unstressed, non-deictic "dummy" pronoun [...] and a NP that semantically functions as the subject of the sentence, the logical subject.' In a comparative perspective, one notices that languages such as Swedish and English do not allow existentials of transitive verbs, whereas Icelandic, German, and Dutch do, as exemplified below in (1):

(1) a. *Det åt en man en pudding. [Swedish]
b. *There ate a man a pudding. [English]
c. Það borðaði maður búðing. [Icelandic]

Faced with the above drastic discrepancy, Platzack ascribes the difference in the use of transitive verbs to the difference in the status of dummy pronouns in the various languages. According to him, if the dummy pronoun is a grammatical subject, as is the case in Swedish and English, then transitive verbs cannot be used. If, on the other hand, the dummy pronoun is not a grammatical subject, as in Icelandic and German, then transitive verbs can occur in existential sentences. Therefore the assumption is that dummy pronouns in Swedish are surface subjects, whereas those in Icelandic are not. Substantiating this point, one notes that, for example, in Swedish, subject-verb inversion is possible, whereas it is not possible in Icelandic. Furthermore, the dummy pronoun of existential sentences in German and Icelandic does not occupy an argument position.

Within the cognitive model, as elaborated in Langacker (1987, 1991), which recognizes a continuum of symbolic units consisting of a semantic and a phonological pole, the dichotomy between syntax and semantics does not play any significant role. Only surface structures are studied, hence there is no justification for transformations of d-structures into s-structures. Equally well, there is no room for syntactic dummies, inserted for purely syntactic purposes.[1] Instead, every element in a sentence needs to be accounted for in semantic terms.

As semantics is believed to be language-specific, the differences found in existential constructions across languages must be ascribed to different factors which shape the construal of an existential scene in a given language. In cognitive grammar one talks of prototypical and marginal category members. If so, then one also recognizes prototypical and non-prototypical PROCESSES.[2] A prototypical process is one which subsumes two PARTICIPANTS, an active one and a passive one, where the former serves as an original source of energy and initiates an interaction, and the latter receives this energy. Langacker (1991: 283) talks of an ACTION CHAIN (see Figure 1 below) which involves one participant transmitting energy to a another, which is thus induced to interact energetically with the next, and so on. The initial energy source is the HEAD of the chain, and the ultimate energy 'sink' constitutes its TAIL.

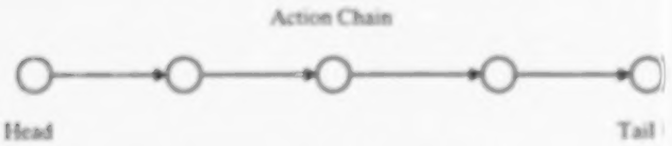


Figure 1 (repeated after Langacker 1991: 283)

The concept of a prototypical process takes its source in the so-called BILLIARD-BALL MODEL, defined in Langacker (1991: 545), that is '[a] fundamental cognitive model that conceives the world as being populated by discrete physical objects that move about and interact energetically when they come into contact.' At the opposite end of the spectrum, one finds processes which depart from the requirements of processual prototypicality. More specifically, one of the participants may be missing from the scene, sometimes even both participants, and the process itself shows the characteristics of an IMPERFECTIVE TEMPORAL RELATION.[3] Keeping in mind the above distinction, we can assume that prototypical existential constructions constitute non-prototypical processes. Typically, an existential process obtains between some kind of SETTING and the entity whose existence, or lack thereof, the sentence announces.[4] The relation between the two evolves in time, hence it is viewed as a process, though there is typically no beginning or end of this relation in the temporal sense.

A characteristic feature of Germanic existentials is that they utilize some kind of semantically vague pronoun sentence-initially, or else, open an existential scene with an adverbial of place, as in the examples in (2) below:

(2) a. Det lå noen bøker på bordet. [Norwegian] (Askedal 1982: 11)
'There lie some books on the-table'.
b. På bordet lå det noen bøker.
'On the-table lie there some books'.
c. Es lagen einige Bücher auf dem Tisch. [German] (Askedal 1982: 11)
'There lie some books on the table'.
d. Auf dem Tisch lagen einige Bücher.
'On the table lie some books'.
e. Det kan finnas ormar i gräset. [Swedish] (Falk 1987: 1)
'There may be snakes in the-grass'.
f. I gräset kan finnas ormar.
'In the-grass may be snakes'.

Contrary to what Askedal says about the semantically empty Norwegian subject *det*, I propose that we analyse it as semantically definable. Its semantic import is that it additionally introduces, to the Norwegian existential sentence, the notion of an ABSTRACT SETTING within which the relation between the adverbial *on the table* and the nominal *some books* unfolds. The German equivalent of the scene which opens with the adverbial of place is not required to be reinforced by the abstract setting *es*. By convention, however, the abstract-setting subject *es* has to open the existential scene when the adverbial is postponed till the end of the sentence. A similar situation to that in the German example takes place in the Swedish sentences. The Norwegian, German, and Swedish equivalents of the adverbials of place *on the table* and *in the grass*, when fronted, play the role of SPATIAL SETTINGS or regions in which the imperfective processes of lying and being unfold. I propose that such fronted adverbials of place function as SETTING-SUBJECTS. Being reified regions, they do not participate actively in the event.[5] They merely provide the concept of some region in which a given entity can be found. In the present framework, it does not seem justified to analyse the sentence-initial pronouns *det* and *es* as entities syntactically different from fronted place adverbials, therefore the former function here as setting-subjects as well. The difference between the pronoun and adverbial subjects appears to be semantic in character, the former being more abstract in their characterization than the latter. Intuitively, one conceives of the region *on the table* more easily than of the one designated by the pronoun *there* or its equivalents in other languages. The region *on the table* can be contrasted with some other region which occupies the area beyond the thing *table*, whereas the region *there*, being virtually unbounded spatially, cannot be so easily contrasted with some other region which does not belong to the scope of *there*.

The fronting of adverbials of place in the Germanic languages is applied much more freely than the fronting of adverbials of time or manner. As observed by Falk (1987: 4), in Swedish it is difficult to find good examples with adverbials of time or manner fronted, as in the examples below:

(3) a. Igår kom fram en flicka till mej och frågade efter vägen.
'Yesterday came up a girl to me and asked the way'.
b. ?Vid middagstid är alltid fullt på restaurangen.
'At dinnertime is always crowded at the-restaurant'.
c. ??Då var livat på Tivoli.
'Then was fun at Tivoli'.
d. *Trots blåsten är varmt.
'Despite the-wind is hot'.
e. *Förskräckligt blåser på kusten.
'Terribly blows on the-coast'.

Evidently, adverbials whose English equivalents are *at dinnertime*, *then*, *despite* and *terribly*, are unfortunate existential openers in Swedish, presumably, not being perceived as reified regions or settings which nicely introduce existential constructions. On the other hand, in a Slavonic language such as Polish, which does not seem to require such an opener, all the sentences in (3) sound perfectly well-formed. However, neither the adverbial setting nor the pronoun setting can be easily extended to the role of an active participant required in the case of some kind of energy transfer. Typically, a spatial region does not perform any visible activity on an entity located within this region. At least, not in any way perceivable for people. Therefore it is hard to conceive of the situations depicted in (4):

(4) a. *On the table smashed a cup.
b. *There kicked a dog.

Now let us consider the crucial semantic difference that obtains between English and Swedish as opposed to Icelandic, on the one hand, and English as opposed to Swedish and Icelandic, on the other. In order to illustrate the point, let us analyse two contexts, repeated after Falk (1989: 47), provided by a transitive and intransitive verb in English, Swedish, and Icelandic in (5):

(5) a. *There digs a woman a grave in the garden.
b. *Det gräver en kvinna en grav i trädgården.
c. Það grefur kona gröf í garðinum.
d. *There danced a man in the garden.
e. Det dansade en man i trädgården.
f. Það dansaði maður í garðinum.

As for the intransitive verb *dance*, it is English in (5d) that does not accept the *there*-subject construction, as opposed to Swedish and Icelandic in (5e) and (5f), respectively. As for the transitive verb *dig*, it is only Icelandic in (5c) that accepts such a scene whereas English and Swedish do not. As evidenced in (5a) and (5d), English does not accept perfectives as well as imperfectives in dynamic use in the *there*-construction.[6] The abstract setting-subject *there* introduces preferably stative processes designating location or existence, *there* being the TRAJECTOR of such a process, and the post-verbal nominal corresponding to the primary LANDMARK.[7] Consequently, the adverbial of place corresponds to the secondary landmark, which undoubtedly carries out the important function of putting in correspondence the conceptual abstractness of *there* to a somewhat more concrete region.

Introduced by the abstract setting-subject *det*, the Swedish existential differs from the English existential in that that the former accepts intransitive processes in dynamic use whereas the latter does not. The Swedish *det* elaborates the trajector of such a process and the setting-subject at the same time. The post-verbal nominal corresponds to the primary landmark and the adverbial of place to the secondary landmark. The important point is that we need to distinguish two levels of analysis here. At the level of a simple clause such as *A man danced in the garden* the thing *man* functions as the subject, hence the trajector of the clause. However, the existential construction, which is conceptually more complex than a simple clause, owing to the presence of the setting-subject, backgrounds the thing *man* and assigns to it the status of the primary landmark. What is highlighted is the subject/trajector of the existential sentence, that is the setting-subject *there*, which assumes the role of the primary element in the sentence.

Moving further on along the axis of scenic complexity, we notice that Icelandic, as in (5c), accepts more complex existential constructions, which become non-prototypical existentials in that that they unfold even transitive processes by means of the introductory setting-subject. As before, the scene has its primary element, that is the trajector-setting-subject. The active participant, which in a simple clause would function as the trajector, acquires the status of the primary landmark. The other participant, that is the patient or theme, corresponds to the secondary landmark, whereas the adverbial of place functions as the ternary landmark. In this way, Icelandic accommodates more elements in an existential construction than the other languages analysed here. However, it needs to be stressed that the primary element of such an existential sentence in Icelandic is the pronoun, which functions as the subject of the overall construction.

Table 1 sums up the kinds of existentials we have considered here.

| trajector setting-subject | process | landmark1 | landmark2 | landmark3 | language |
|---|---|---|---|---|---|
| Það | borðaði | maður | búðing | ------------- | Icelandic |
| Það | grefur | kona | gröf | í garðinum | Icelandic |
| Det | lå | noen bøker | på bordet | ------------- | Norwegian |
| På bordet | lå | det | noen bøker | ------------- | Norwegian |
| Es | lagen | einige Bücher | auf dem Tisch | ------------- | German |
| Auf dem Tisch | lagen | einige Bücher | ------------- | ------------- | German |
| Det | kan finnas | ormar | i gräset | ------------- | Swedish |
| I gräset | kan finnas | ormar | ------------- | ------------- | Swedish |

Table 1

The main point this paper has raised is that existential constructions, in the languages we have considered, differ in the construal of the scene. More precisely, it is the arrangement of the elements of the scene as well as different access paths to these elements that cause altered construals of essentially the same situations.

Tentatively, I assume that a language such as Icelandic introduces existential scenes by means of the pronoun *það* despite the kind of process that unfolds in this scene. Therefore there may be just some activity performed by the primary landmark during which no energy interaction between two participants obtains, like in (5f), or some interaction may take place between three landmarks located within the abstract region, as in (5c). English existentials, as in (5a) and (5d), do not seem to allow dynamic activities altogether, showing preference for stative processes. Somewhere in between, Swedish existentials accept dynamic processes which do not entail energy transfer between participants, as in (5e). However, such distinctions, being part of language convention, should not influence our judgements about the allegedly different syntactic statuses of the essentially identical elements in related languages. In view of the above, I conclude that assigning the role of the subject to those equivalents of *there* which allow only intransitive processes, and not transitive ones, is selective at best, if not totally arbitrary.

*Department of English*
*Maria Curie-Sklodowska University*
*pl. M. C. Sklodowskiej 4*
*20-031 Lublin, Poland*
*e-mail: pettar@klio.umcs.lublin.pl*

**Notes**

* Work on this paper was partially carried out in the English Department, Umeå University, Sweden, thanks to a research grant from the Swedish Institute, Stockholm, which I gratefully acknowledge.

[1] Semantically or phonologically empty entities are ruled out in this framework by the so-called CONTENT REQUIREMENT. For details, see, for example, Langacker (1991: 546).

[2] Process is a technical term in cognitive grammar and covers such categories as verbs, clauses, and sentences. In cognitive terms a process is a relation which has a positive temporal profile, as opposed to an atemporal relation, that is an adjective, adverb, and preposition.

[3] An imperfective temporal relation, or an imperfective process, profiles a relation as unchanging through time, as opposed to a perfective process.

[4] Langacker (1991: 553) defines setting as '[a] global, inclusive region within which an event unfolds or a situation obtains.' The concept of setting is typically juxtaposed with the one of participant.

[5] By a reified region is meant here the situation in which the relational profile of an otherwise prepositional phrase is shifted to the profile of some region, that is a kind of THING. Cases such as *Near the fire is warmer* or *Under the bed is all dusty* are discussed in more detail, for instance, in Langacker (1995: 72-3).

[6] For an in-depth analysis of *there*-constructions in English, see Case Study 3 in Lakoff (1987).

[7] Trajector and landmark are two terms, postulated by Langacker, for the designation of the primary and secondary figures, respectively, within a profiled relation. Additionally, one also recognizes primary and secondary landmarks. A ternary landmark is postulated in the present paper.

## References

Askedal, John Ole. 1982. On the syntactic representation of so-called 'existential-presentative sentences' in Norwegian and German. A contrastive analysis, ed. by Thorstein Fretheim and Lars Hellan, 11-25. Trondheim: Tapir.

Clark, Eve V. 1978. Locationals: existential, locative, and possessive constructions, ed. by Joseph H. Greenberg, 85-126. Stanford: Stanford University Press.

Falk, Cecilia. 1987. Subjectless cluases in Swedish. (Working papers in Scandinavian syntax, 32.) Lund: Department of Scandinavian Languages.

Falk, Cecilia. 1989. On the existential construction in the Germanic languages. (Working papers in Scandinavian syntax, 44.) Lund: Department of Scandinavian Languages.

Falk, Cecilia. 1993. Non-referential subjects in the history of Swedish. Lund: Department of Scandinavian Languages dissertation.

Freeze, Ray. 1992. Existentials and other locatives. Language 68.553-95.

Holmes, Philip, and Ian Hinchliffe. 1994. Swedish. A comprehensive grammar. London: Routledge.

Lakoff, George. 1987. Women, fire, and dangerous things. Chicago: The University of Chicago Press.

Langacker, Ronald W. 1987. Foundations of cognitive grammar, vol. 1, Theoretical prerequisites. Stanford: Stanford University Press.

Langacker, Ronald W. 1991. Foundations of cognitive grammar, vol. 2., Descriptive application. Stanford: Stanford University Press.

Langacker, Ronald W. 1995. Possession and possessive constructions, ed. by John R. Taylor and Robert E. MacLaury, 51-79. Berlin: Mouton de Gruyter.

Lyons, John. 1967. A note on possessive, existential and locative sentences. Foundations of Language 3.390-6.

Maling, Joan. 1987. Existential sentences in Swedish and Icelandic: Reference to thematic roles. (Working papers in Scandinavian Syntax, 28.) Lund: Department of Scandinavian Languages.

Platzack, Christer. 1983. Existential sentences in English, German, Icelandic and Swedish, ed. by Fred Karlsson, 80-100. Helsinki: University of Helsinki.

# WORD OF MOUTH

Communication is power. An idea, passed from person to person, and village to village, can transform the world.

Start with the right idea.

**Linguistics and Language Behavior Abstracts**
offering

- abstracts of scholarly articles and books
- bibliographic entries for subject specific dissertations and book and other media reviews.

LLBA's timely and comprehensive coverage speaks volumes on current ideas in linguistics and language research.

Available in print or electronically through the Internet Database Service (*www.csa.com*). Contact *sales@csa.com* for a trial Internet access or a sample issue.

**LLBa**

***Linguistics and Language Behavior Abstracts***
*Published by Cambridge Scientific Abstracts*
7200 Wisconsin Avenue • Bethesda, Maryland 20814 • USA
+1 301-961-6700 • Fax: +1 301-961-6720 • E-mail: *sales@csa.com*

# Odense Working Papers in Language and Communication

*OWPLC* publishes scholarly articles written by members of the Institute of Language and Communication at the University of Southern Denmark, Main Campus: Odense University and their Danish or foreign colleagues.

Nos 13-19 include:

No. 13 Lene Schøsler & Karen Van Durme:
*The Odense Valency Dictionary*

No. 14 Karl-Heinz Pogner (ed.):
*Writing: Text and interaction*

No. 15 Karin Van Durme (ed.):
*The valency of nouns*

No. 16 Dorthe Bleses & Johannes Wagner (eds.):
*Papers in First Language Acquisition*

No. 17 Dorthe Bleses:
*The role of input, productivity and transparency in Danish children's acquisition of past tense morphology*

No. 18 Teresa Cadierno (ed.):
*L2 Listening comprehension*

No. 19 Carl-Erik Lindberg & Steffen Nordahl Lund (eds.):
*17th Scandinavian Conference of Linguistics, vol I-II*
*20-22 August 1998*

Copies can be obtained from

**Institute of Language and Communication**
**University of Southern Denmark,**
**Main Campus: Odense University**
Campusvej 55
DK-5230 Odense M.
Denmark

General Editor: Steffen Nordahl Lund
(E-mail: snl@language.sdu.dk.)

***Odense Working Papers in Language and Communication*** publishes scholarly articles written by members of the Institute of Language and Communication at SDU - Odense University as well as papers by their Danish or foreign colleagues.

This issue of ***Odense Working Papers in Language and Communication*** contains most of the papers presented at the 17th Scandinavian Conference of Linguistics, which was organised by the Institute of Language and Communication, August 20-22, 1998.

Some 80 linguistics from Scandinavia and other parts of Europe, as well as from the USA and New Zealand, took part in the conference, which adressed the following main topics: The Use of IT in Grammatical Analysis/Parsing; Functional/Cognitive Linguistics; Language in Interaction; Endangered Languages.

UNIVERSITY OF SOUTHERN DENMARK

279

UNIVERSITY OF SOUTHERN DENMARK

# 17th Scandinavian Conference of Linguistics, vol. II

**Carl-Erik Lindberg**
**& Steffen Nordahl Lund (eds.)**

No. 19
April
2000

ODENSE
WORKING PAPERS
IN LANGUAGE
AND COMMUNICATION

280

**Carl-Erik Lindberg & Steffen Nordahl Lund (eds.)**
**17th Scandinavian Conference of Linguistics, vol. II**

Odense Working Papers in Language and Communication
No 19, vol II, April 2000

General editor: Steffen Nordahl Lund (snl@language.sdu.dk)

Institute of Language and Communication
University of Southern Denmark,
Main Campus: Odense University
Campusvej 55
DK-5230 Odense M.
(lfn@language.sdu.dk)

Printed in Denmark by Odense Universitets Trykkeri

# Indhold

Forord i

## Language in interaction

*Elin Fredsted*
Om konflikt og kooperation i samtaler 1

*Karsten Gramkow*
Overlap Management and Interactional Competence 17

*Annette Grindsted*
Stereotypes in action 33

*Birgitte Jacobsen*
Danish in Greenland – a new dialect?
Preliminary results from a pilot project 51

*Helen Korsgaard Maxwell*
Meanings of the Word "Work" in a Danish Unemployment Office 63

*Hans Jørgen Ladegaard*
Grices kooperative principper i kritisk belysning:
Non-kooperation som diskursiv strategi i asymmetriske diskurser 75

*Leila Naseh*
Codeswitching Between Persian and Swedish 89

*Frederikke Trondhjem*
En sprogundersøgelse af en såkaldt halvsproget person 111

## Endangered Languages

*Ida Elisabeth Mørch*
How fast will a langauge die when it is officially no longer spoken? 125

*Marja Leinonen*
Komi, an endangered language? 135

*Ida Toivonen*
Phonetic salience and language change: Verb paradigms in Inari Sami 151

*Jan Heegård*
Linguistic and political aspects of alphabet-making for a
threatened language 161

## Forord

Institut for Sprog og Kommunikation, Odense Universitet, arrangerede i dagene 20.-22. august 1998 det 17. Nordiske Lingvistmøde på Hotel Nyborg Strand ved Storebælt.

Denne række lingvistmøder arrangeres skiftevis af de enkelte nordiske lande med ca. 1 1/2 års mellemrum, og har gennem årene opnået nærmest institutionslignende status. Møderne er åbne for foredrag inden for samtlige lingvistiske discipliner; dog har arrangørerne mulighed for at udvælge og prioritere et antal hovedtemaer. Til lingvistmødet 1998 havde de lokale arrangører prioriteret følgende emner:

- The Use of IT in Grammatical Analysis/Parsing
- Functional/Cognitive Linguistics
- Language in Interaction
- Endangered Languages

Herudover var der en åben sektion med andre temaer.

I mødet deltog følgende hovedtalere, der alle er internationalt førende forskere inden for deres område: professor Östen Dahl, Stockholms Universitet, Professor John Taylor, University of Otago, New Zealand, professor Fred Karlsson, Helsingfors Universitet, professor Jürgen Streeck, University of Texas at Austin og professor Erling Wande, Stockholms Universitet.

I dette års lingvistmøde deltog ca. 80 sprogforskere, først og fremmest fra Skandinavien, men også fra det øvrige Europa, fra USA og New Zealand.

Det var på forslag fra præsidenten for the Nordic Association of Linguists, Helge Dyvik, Universitetet i Bergen, at området Truede Sprog blev optaget i dette års program. Området omfatter mindre sprog og især sprog som af historiske, sociale eller politiske grunde er ved at ophøre med at eksistere. I den "sprogudjævning" som i disse år foregår over hele verden, ikke alene ved udligning mellem rigssprog og dialekter, men også ved bl. a. "majorisering", hvor enkelte mindre sprog eller minoritetssprog trues af tilsigtet eller utilsigtet udryddelse, er en lingvistik og tillige kulturvidenskabelig beskæftigelse med dette område væsentlig, såvel ud fra et sprogligt og kulturelt, som ud fra et samfundsmæssigt og demokratisk synspunkt.

De fleste af deltagerne i 1998 årsmødet bidrog med foredrag over emner inden for de fire hovedtemaer. En række af disse foredrag udgives herved i serien *Odense Working Papers in Language and Communication* (hovedtalernes bidrag vil blive publiceret andetsteds).

# LANGUAGE IN INTERACTION

# Om konflikt og kooperation i samtaler

*Elin Fredsted*

**Abstract**

The present paper deals with dialogues in tourist information centres between Danish and German native and foreign language speakers.

An analysis of 120 dialogues shows that in conversations at tourist offices especially Danes tend to use direct and informal language without modifiers or politeness marking linguistic elements.

Native German speakers, however, are characterized by preferring more elaborated dialogues with conventional politeness.

More specifically, a close analysis of a dialogue containing special features of conflict is presented. The analysis focusses on an investigation of the employee's avoidance strategies in this situation. It is shown that both parties do their best to hide a deficit of cooperation under a surface of benevolence and politeness.

## Beskrivelse af korpus

Genstanden for min undersøgelse er den eksterne kommunikation inden for turistbranchen. Anledningen var det EU-støttede projekt "Dansk og tysk som fremmedsprog inden for handel og erhverv i Schleswig-Holstein og Sønderjyllands Amt", som Karen Margrethe Pedersen og jeg arbejdede med i årene 1992-1994. Vores opgave i dette projekt var bl.a. at udvikle curricula og undervisningsmaterialer til fremmedsprogsundervisning i dansk og tysk for ansatte inden for turistbranchen.

Undervisningsmaterialet blev baseret på observation og registrering af den kommunikation på modersmålet og på fremmedsprog, der fandt sted på fire turistkontorer i Sønderjylland og Slesvig-Holsten. Vi har observeret og nedskrevet samtaler på stedet, men også deltaget mere aktivt i forskellige arbejdsprocesser. I de senere faser har vi så optaget lydbånd og videofilm på de steder, hvor vi allerede havde lært arbejdsgangen at kende. Resultaterne heraf udgjorde korpus for de sprogkurser for ansatte inden for turistbranchen, som vi udviklede i forlængelse af undersøgelsen. Det sprogkorpus, vi havde samlet, gjorde jeg så dernæst til genstand for en sprogbrugsanalyse.

Under optagelserne på turistkontorerne i Sønderjylland og Slesvig-Holsten gjorde vi en række umiddelbare sproglige iagttagelser, som var meget påfaldende. Et eksempel: Når danskere går ind på et turistkontor for at bede om information, stiller de typisk et ganske direkte spørgsmål uden høflige omsvøb og får et lige så direkte svar fra medarbejderen.

Turist: Hvad kan man egentlig se her?
Medarbejder: Her er en brochure med informationer om, hvad der er at se.

De tysktalende turister tenderer derimod til mere langtrukne, konventionelt høflige udtryk som:

Turist: Ich bin das erste Mal hier. Das sieht ja alles toll aus da draußen. Was kann ich mir denn hier sonst so ansehen?
Medarbejder: Ja, da kann ich Ihnen vielleicht mal eine breite Palette Prospekte zusammenstellen und Ihnen mal zeigen, was es so zu sehen gibt.
Turist: Sehr nett!

En af iagttagelserne var, at sproget (dansk eller tysk) var knyttet til en kulturel kode, som mest påfaldende kom til udtryk i forskellige grader af sproglig høflighed og i forskellige grader af sproglig elaboration. Vi var i første omgang selv forbløffede over denne tydelige forskel i sprogligt udtryk, som til gengæld får danske turister og turistmedarbejdere til at virke temmelig direkte, når de taler tysk som fremmedsprog, fordi de virker så korte for hovedet. Det viste sig da også, at de danske turistmedarbejdere i nogle tilfælde langt fra er 'høflige' nok, når de informerer tyske turister på tysk. De lever med andre ord ikke op til den grad af formel sproglig høflighed og elaboration, som den tysksprogede turist kan formodes at forvente. Begge sprog og kulturer har bestemte forestillinger om sproglig høflighed, men disse forestillinger er ikke identiske, og det er den grundlæggende interesse i min ph.d.-afhandling "Analyser af dansk og tysk talesprog" at analysere de sproglige udtryk og påpege konkrete ligheder og forskelle (Fredsted 1998).

Jeg har analyseret i alt 120 samtaler nord og syd for grænsen og fundet frem til, at der er tale om en meget høj grad af konventionelle handlingsforløb, en høj grad af standardisering og gentagne strukturer i samtalerne, så der kan opstilles en slags typologi for informationssamtaler på turistkontorer: Turisten stiller et eller to spørgsmål eller fremsætter en anmodning. Turistmedarbejderen besvarer spørgsmålet, efterkommer anmodningen (f.eks. om kort og brochurer). Derefter siger man så måske tak og farvel:

| | | | |
|---|---|---|---|
| initial fase | ➡ | 1. dialogåbningssignal - hilsen | |
| | ➡ | 2. fatiske interaktionssignaler | |
| | ➡ | 3. introduktion af tema | |
| medial fase | | "forhandling" | ➡ |
| | | ↕ | |
| | | (introduktion af nyt tema | ➡) |
| final fase | | 1. konklusion | ➡ |
| | | 2. tak | ➡ |
| | | 3. afskedshilsen | ➡ |

Skemaet skal læses på følgende måde: Undersøgelsen af samtaler på turistkontorerne har vist, at en samtale kan åbnes med en initialfase, der kan bestå af en hilsen eller en anmodning/et spørgsmål, der indeholder høflighedsmarkører af en eller anden art (1. og 2.). En anden mulighed er at åbne en samtale ved direkte at introducere det tema, som samtalen skal dreje sig om (3.). Samtalepartnerne har mulighed for at gå ud af samtalen, når spørgsmålet er besvaret/ anmodningen efterkommet (første pil på højre side), men der er også den mulighed at stille et nyt spørgsmål/ fremsætte en ny anmodning. Herefter kan samtalepartnerne enten gå hver til sit eller slutte med en konklusion, en tak og måske også en afskedshilsen. Skemaet er en idealtype, som samtalerne på de to sprog forholder sig til på en varierende måde: De danske samtaler har en tendens til at begynde i den initiale fases trin nr. 3 og til at afslutte med en konklusion og eventuelt en tak i finalfasen. D.v.s. at de danske samtaler tendentielt koncentrerer sig omkring den tematisk orienterede medialfase. De tysksprogede samtaler (med tyske turister og tyske medarbejdere) har derimod i højere grad en tendens til at begynde i den initiale fases trin 2 og slutte med en "tak" (den finale fases 2. trin). Kort formuleret betyder det, at de tyske samtaler i indlednings- og afslutningsfasen i mindre grad er tematisk orienterede, mere fatiske og ritualiserede end de danske.

Den samtale, jeg gerne vil præsentere i det følgende, er et resultat af vores videooptagelser og er på nogle områder atypisk, netop fordi det gennemgående mønster, der er beskrevet ovenfor, brydes. Først og fremmest er samtalen atypisk, fordi den er væsentlig længere end de fleste andre, og fordi den - i modsætning til flertallet af de samtaler, vi har registreret - indeholder et konfliktpotentiale, der viser, hvordan en turistmedarbejder reagerer i en kritisk situation, der bevæger sig ud over den almindelige kommunikative rutine.

## 'Wissen Sie, was wir gut fänden'

Omstændighederne omkring den følgende samtale kræver en nærmere forklaring. Under optagelserne i byen F. kom en sortklædt herre på ca. 70 - 75 år med hvidt hår og hvidt skæg ind på turistkontoret ledsaget af en yngre, elegant klædt dame. Det følgende er en udskrift af den samtale, der udspandt sig mellem den ældre herre (R) og de tilstedeværende på turistkontoret:

R: den sortklædte herre

M: Medarbejderen

T: en anden turist, dernetop er i færd med at leje et værelse, da R træder ind.

For at forstå samtalens kontekst (og Rs forveksling af T med en kendt person) er det vigtigt at vide, at tysk TV på samme tid var i færd med at optage scener til en spillefilm på gaden et par hundrede meter fra turistkontoret. Hvad der ikke fremgår af den efterfølgende transskription er, at R – skønt han taler næsten korrekt tysk – efter min vurdering ikke har tysk som førstesprog. Dette giver sig udslag i et usædvanligt, litterært ordvalg *(wohlfeil)*, hyperkorrekte former (*dieses*, recte: *dies*) og til dels også i en udtale med ikke-tysk accent, der gjorde, at det ind imellem var vanskeligt for de tilstedeværende at forstå, hvad han sagde.

R: tag
M: guten tag!
R: irgendwie prominent (------)?
((Afbrydelse: 2 telefonsamtaler på ca. 25 sekunder og 62 sekunder. M henvender sig derefter til T:))
M: klein' moment, bitte mal. darf ich zuerst die herrschaften?
T: ja.
R: wir sind zum- das ist fast videoreif,
M: ja.
R: =zum 90. geburtstag von Hermann Hansen unserem ehrenmitglied aus Hamburg hier.
M: ja ja! onkel Hermann!
R: ich bin der vorsitzende und begründer der deutsch-jüdischen gesellschaft in Hamburg.
und ich wollte für unsere dritte vorsitzende, die zum erstenmal hier ist,
M: =schön!
R: einiges an material von hier mitnehmen.
M: gerne!
R: wenn Sie ein bisschen zusammenstellen können.
M: prima!
R: Sie kennen es ja schon,
M: ja gerne!
R: wenn Sie das zusammenstellen können, wäre das sehr schön.
M: ja, mache ich gerne! ich mache Ihnen eine palette fertig.
R: ja, das wäre sehr schön, wenn Sie uns eine palette fertig machen können, ja.
M: schon fix und fertig gepackt. ((ler))
R: oh!
M: sehen Sie mal. da lege ich auch noch ein stadtprospekt mit rein, und ich würde auch
sagen einen rundgang durch die stadt.

R: gerne.
M: ja?
(4)
R: ja. wissen Sie was wir gut fänden.
M: ja?
R: wir haben hier als allererstes eine aktivität gemacht. das wissen Sie wahrscheinlich,
((R vender sig mod T og peger på ham))
%Sie kommen mir etwas bekannt vor, aber ich bin mir nicht sicher%
es wurde vor zehn jahren hier ein synagogenstein getextet,
M: ja,
R: =und der steht ja da auch,
M: ja.
R: genau original unser text. und dann ist auch für die deportierten von
Friedrichstadt, die deportierten juden ein stein hingekommen, den wir
auch initiiert haben. wir würden das sehr begrüßen, wenn das auch
fremdenverkehrsmäßig- natürlich keine attraktion
[in dem sinne]
M: [ja]
R: =und so weiter nicht?
aber wenn das dann auch irgendwo mal erwähnt würde,
M: ja:
R. es gab hier juden und so weiter und so fort. und es gibt hier ein' stein.
M: gerne
R: das ist ja eine sehenswürdigkeit.
M: ja:
R: das ist nämlich das originalgebäude die schule und die synagoge
M: die [synagoge ja.]
R: [das ist auch einzigartig,] ja? das synogogengebäude.
M: ja.
((R vender sig om mod os og kameraet))
R: ehm: Sie wussten dieses? Sie wissen dieses. und das halte ich auch für sehr wichtig.
M: ganz toll, ja
R: machen Sie das fürs fernsehen?
EF %nein für die schule%
R: für die schule. ich bin lehrer, ja? machen Sie das für F.?
EF nein, für Dänemark,% für den deutschunterricht in Dänemark.%
R: für Dänemark. ja! ganz wichtig! die dänischen juden gerettet worden, ja? ist Ihnen bekannt
alles?
EF ja.
R: wunderbar, so.
M: gut. ich gebe das gerne weiter. da ist die palette, bitte sehr!
R: was sind wir Ihnen schuldig?
M: gar nichts. [das ist in ordnung.]
R: [wunderbar, wunderbar.]
das darf ich mal mitnehmen?
M: ja, ja.
R: haben Sie sonst irgendetwas neueres?
M: das ist alles in der (.) palette mit drin.

73 R: über die 13 religionsgemeinschaften, so viele sind es ja mal gewesen, nicht? wenn ich 74
75 M: ((til kvinden)) % [das] ist ganz neu.%
76 R: =haben Sie da irgend etwas (.) [kulturgeschichtliches]?
77 M: [nein, da habe] ich momentan nichts. da würde ich empfehlen, dass Sie da
78 vielleicht doch noch mal bei der buchhandlung Kruse nachschauen.
79 R: ah ja! die stadt macht aber
80 nichts weiteres?
81 M:wir haben nur dieses touristsche material hier
82 R: ah ja.(.)
83 ich sehe da was, was ich noch nicht gesehen habe dithmarscher museumswerkstätte?
84 M: das ist in Meldorf (4) %meine alte heimatstadt.%
85 R: und der veranstaltungskalender für 94, wenn Sie mir persönlich das auch noch mal geben
könnten?
86 M: um:, ja. das machen wir. ich mach eine kopie dazu. das ist noch ein bisschen was ergänzend
87 R: gern.
88 M: so:. ich hoffe, daß ich Ihnen damit helfen konnte
89 R: ja. vielen dank.
91 M: das ist frei ja. selbstverständlich.
92 R: ist das in der palette drin gewesen?
93 M: nein.
94 R: ((til kvinden)) nehmen Sie mal, Sie sind ja auf vögel (.) spezialisiert.
95 M: gut.
96 R: vielen dank, auf wiedersehen.
97 M: wiedersehen ich wünsch Ihnen einen schönen tag noch! tschüss
98 R: ((til os)) und Ihnen alles gute!wie sagt man auf dänisch? auf wiedersehen?
99 EF [farvel]
100 M: [farvel]
101 R: farvel. das muss man lernen, nicht? wir waren in Hamburg ja auch von Dänemark
umschlossen,
102 Sie wissen das? Altona war dänisch, Wandsbek war dänisch. hat für die hamburger juden
103 eine ganz große bedeutung gehabt. schwärmen wir in der ganzen Bundesrepublik von.
104 EF es war ja- Altona war
105 [ja tolerant.]
106 R: [war tolerant.] Hamburg war nicht tolerant. und wir sagt man auf dänisch wiedersehen?
107 EF farvel
108 R: farvel, farvel, goodbye. alles gute.
109 EF danke gleichfalls!
((R stiller sig op foran kameraet og bukker))

R kommer ind på turistkontoret og ser straks, at der foregår videooptagelser. Hans ledsagerske ser også kameraet, men holder sig derefter diskret i baggrunden og ytrer ikke et ord, før hun igen er trådt ud af turistkontoret. Mens M telefonerer, tager R en af Ts brochurer (brochuren om F.) op fra disken og bladrer den igennem, hvorpå T samler resten af sine brochurer i en bunke foran sig. Da M henvender sig til R,

fortæller han, hvorfor han er i F., præsenterer sig selv og sin ledsagerske, på hvis vegne R beder om materialer. Samtalen begynder m.a.o. som en rutinesag: anmodning om og udlevering af brochurer.

Det forekommer ganske vist ikke så hyppigt i de indsamlede data, at personerne præsenterer sig selv. Undtagelser herfra er dog telefonsamtaler og samtaler, der (som denne) er optaget på video: I telefonsamtalerne præsenterer folk sig altid rutinemæssigt med navn; i syv af de 24 videooptagelser sker der en form for selvpræsentation uden navns nævnelse, uden at dette vel at mærke synes at dreje samtalerne i en bestemt retning. Det er åbenbart den synlige tilstedeværelse af et kamera, der fremkalder en slags selviscenesættelse hos turisterne.

Det første egentlige brud med den ovenfor beskrevne kommunikative rutine opstår i forbindelse med den korte pause, (på 4 sekunder), der opstår før Rs ytring: 'ja, wissen Sie was wir gut fänden' (linje 29-30). Pausen indicerer, at R har et specielt ærinde. R fremsætter herefter en anmodning, hvis baggrund kræver en nærmere forklaring: R har under gennembladringen af Ts brochure om F. fastslået, at en mindesten ved synagogen (og dermed om det jødiske samfund i F., der blev udryddet under nazismen), en mindesten om de deporterede jøder og selve synagoge- og skolebygningen ikke er omtalt nogen steder i turistbrochuren. Dette påtaler R nu, men indleder denne påtale med en positiv formulering: 'ja, wissen Sie was wir gut fänden' (linje 30).

Denne ytring danner indledning til et brud i samtalen. M har indtil dette tidspunkt opfattet R som en almindelig informationssøgende turist, og intet i samtalen tyder på, at samtalepartnerne indtil da har orienteret sig i forhold til et problem- eller konfliktpotentiale. [1] Rs problem er, at der ikke er tale om nogen kohærens eller naturlig overgang. Overgangen er først og fremmest vanskelig, fordi den bryder aktivitetstypen 'orienterende samtale', og fordi Rs ytring betegner et rolleskift i samtalen fra informationssøgende turist til repræsentant for en minoritet. Det bør understreges, at R i sin påtale (linje 30-54) ikke på nogen måde er uhøflig eller anklagende: Som anført ovenfor indleder han med en positiv formulering (linje 30), men også i det følgende anvender han talrige forsigtige og høflige modifikatorer som f.eks. modificerende partikler ('aber wenn das *dann auch irgendwo mal* erwähnt würde' (linje 45)), brug af konjunktiv II (linje 30, 40, 45), konditionale udtryk (linje 45) og koncessessive udtryk (linje 41-42), der signalerer, at R ikke intenderer nogen form for konfrontation, men udtrykker sig bevidst forsigtigt og appellerende (jf. linje 44: '=und so weiter nicht?' og linje 53: 'ja?' med stigende intonation).

Ms talebidrag indskrænker sig i denne fase af samtalen (linje 30-54) til rent fatiske *back- channel*-signaler: 'ja' med faldende intonation (linje 36, 38, 43, 46, 50, 54) og ja-ækvivalerende udtryk som 'gerne' (linje 48) og ekko som 'die Synagoge ja' (linje 52). Signifikant er, at han skifter fra 'ja' med stigende intonation i linje 28 og 31 til 'ja' med faldende intonation, et skift, der er med til at markere et faldende engagement i samtalen fra Ms side. På overfladen er M stadig høflig og kooperativ, idet han udelukkende udsender positive sproglige signaler, men indholdsmæssigt er hans reaktion så intetsigende, at R formodentlig gennemskuer, at den dækker over en de facto afvisning. Nærmest absurd bliver samtalen, hvor R forsøger at skitsere en kort tekst til turistbrochuren, som M responderer på med 'gerne' (linje 48).

Da R allersenest i linje 55 mærker, at han ingen egentlig respons finder hos M, vender han sig mod kameraet og os for at åbne en sidehandling og søge respons hos os som tilhørere. R forventer sig tydelig nok en slags publikumseffekt på grund af kameraet (jf. også Rs henvendelse til T, hvor R formoder, at T er en kendt personlighed i linje 3 og 34). Denne sidehandling afsluttes i linje 65, hvor M overrækker konvolutten med brochurerne til kvinden med taleturen:

M: gut. ich gebe das gerne weiter. da ist die pallette, bitte sehr! (linje 65)

Fra dette punkt i samtalen forsøger M gentagne gange at bringe samtalen ind i en afsluttende fase ved hjælp af terminalmarkører udtalt med stærkt faldende intonation:

M: [das ist in ordnung]. (linje 67)
M: so:. ich hoffe daß ich Ihnen damit helfen konnte. (linje 88)
M: gut. (linje 95)

I den første ytring (linje 65) forsøger M at lukke begge temaer på én gang: påtalen ('gut. ich gebe das gerne weiter') og udleveringen af informationsmaterialet ('da ist die palette'). De tre efterfølgende ytringer relaterer sig derimod udelukkende til temaet informationsmateriale. Ms bestræbelser på at føre samtalen ind i en afsluttende fase kommer også til udtryk ved, at M referer bagud i samtalen til allerede udleveret materiale (linje 72), henviser til andre steder, hvor R kan henvende sig (linje 77-78), og ved at fralægge sig kompetence (linje 65 og 81). Til gengæld forsøger R gentagne gange – og derfor efterhånden på en ret påfaldende og vedholdende måde – at genåbne samtalen med M ved at stille nye spørgsmål og anmodninger, der alle relaterer sig til temaet 'informationsmateriale' (i alt 8 gange i linje 69, 71, 73-76, 79-80, 83, 85, 90, 92). Dog forsøger R også ad denne vej implicit at bringe de kultur- og dermed religionshistoriske forhold på banen igen, idet han beder om kulturhistorisk materiale (73-76, 79-80) for dermed at skabe affinitet til sin påtale af fortielsen af byens jødiske fortid. M bestræber sig på at lukke temaet ved at henvise til boghandelen og ved at insistere på, at R kun kan få rent turistisk materiale. Pointen er, at hele samtaleforløbet fra 'das darf ich mal mitnehmen' (linje 69) kan ses som en lang kamp mellem M og R om at lukke hhv. opretholde samtalen.

At det imidlertid er lykkedes for M at blokere temaet 'jødeforfølgelser' fremgår af det faktum, at R nu kun tager dette tema op over for kameraet og os (linje 101 ff): Det er altså åbenbart lykkedes for M at spalte samtalen, så R må anvende et adressatmønster, der hedder: M som adressat for ytringer vedr. informationsmaterialet, kameraet og os som adressat for omtale af det jødiske tema. R afslutter da også sin optræden med et lille foredrag henvendt til os og kameraet om de danske jøders redning og Altonas betydning for de tyske jøder i Hamburg .

M fastholder sin del af samtalen med R som værende monotematisk: M demonstrerer, at han kun accepterer ét tema som genstand for samtalen. Ms prœterminale markører kan altså kun forstås som et forsøg på at bringe samtalen ind i en afsluttende fase. R bestrœbelser på at opretholde samtalen sker da heller ikke med introduktion af nyt tema, men ved at fastholde allerede omtalte temaer. At det principielt er muligt at genåbne en samtale på et hvilket som helst stadium af en afslutningssekvens demonstrerer R meget klart ved at tematisere selve afskedshilsenen (linje 98 ff.).

At R søger publikum er måske ikke så bemœrkelsesværdigt, da han (fra linje 30) definerer sin rolle i samtalen som reprœsentant for en interessegruppe, en religiøs-etnisk minoritet. En mere uddybende kommentar krœver derimod Ms reaktioner: Hvorfor vil M med alle (sproglige) midler undvige?

## Aktivitetstype og roller

Jeg vil i det følgende prøve at forklare Ms sproglige interaktionsmønster ud fra to begreber, der har med kontekstualitet at gøre:
1) det første begreb vil jeg kalde *activity framework* med en venlig hilsen til Wittgensteins begreb om 'Sprachspiele' (Wittgenstein 1953) og Levinsons artikel om 'Activity types and language' (Levinson 1979).
2) det andet begreb er *participation framework*; hermed betegner jeg deltagernes rollemønstre og interaktionsmønstre.
Pointen er, at i denne samtale spilles de to typer af *framework* ud mod hinanden.

1) Aktivitetstype / *activity framework*:

Ganske tydeligt hersker der forskellige opfattelser af, hvilket 'sprog-spil', der kan og bør spilles i den givne situation. Ikke kun Ms verbale udtryk, men også den non-verbale kommunikation på videooptagelsen viser tydeligt, at samtalen tager en retning, som M finder problematisk: Ved omtalen af de deporterede jøder (linje 39) ændres Ms ansigtsudtryk markant, idet der fra dette tidspunkt af samtalen træder to skarpe rynker frem mellem hans øjenbryn. At samtalen her forlader det turistiske områder og glider over i et tabubelagt politisk tema, fremkalder en negativ non-verbal reaktion, der genspejles i Ms mimik og i et 'stivnet' kropssprog, mens de verbale reaktioner stadig er formelt set høflige og postive, men - som påpeget ovenfor - tydeligt rettet mod at kommunikere, at han ikke er den rigtige adressat for Rs påtale (jf.: 'gut. ich gebe das gerne weiter.' (linje 65)). Alle Ms talebidrag fra linje 65 kan ses under én synsvinkel: at bringe samtalen tilbage til det sprog-spil, der i Ms forståelse af situationen skal spilles her. Det gør han ved

a) at respondere pro forma på alle andre temaer end dem, der har med turisme at gøre med overfladiske, høflige, men intetsigende responser,

b) aktivt at lede samtalen tilbage på turistiske temaer og genstande, der hører hjemme inden for den *activity framework*, som M skal fungere indenfor: turistkontorets verden, besvarelse af turistiske spørgsmål, udlevering af brochurer, dvs. verbal og non-verbal respons på én bestemt type af spørgsmål eller statements,

c) at forsøge at bringe samtalen ind i en afsluttende fase.

Formuleret lidt anderledes: M definerer sin ramme til udelukkende at være den professionelle. Han er som turistmedarbejder ikke forpligtet som borger eller privatmand til at komme med statements, der ikke har noget med udførelsen af hans profession at gøre. En vigtig del af hans professionelle rolle er til gengæld at undgå konflikter, og dem undgår han ved at give høfligt imødekommende, men intetsigende respons.

R har derimod en interesse i at undgå, at samtalen bryder sammen og samtidig at få lejlighed til at udfolde sit tema. Da han drager de rigtige inferenser af Ms tilbageholdenhed, er hans første strategi at få sit tema til at passe ind i Ms *activity framework* (f.eks. med omtalen af synagogebygningen som en turistattraktion i linje 49-53). Da det heller ikke umiddelbart lykkes (p.g.a. Ms intetsigende ekko-svar linje 52) søger han at fastholde sit tema, men skifte sin adressat ud, idet han henvender sig til tilhørerne og søger at inddrage dem i samtalen, altså skifte *participation framework*. Hans anden strategi for at fastholde temaet bliver altså at udvide *participation framework* ved at inddrage tilhørerne i samtalen (fra linje 55).

## Sammenfatning

Der sker en spaltning af diskursen: Diskursen mellem R og M fortsætter på overfladen, men under den kooperative overflade har M for længst koblet fra. R gennemskuer det, men spiller spillet med og vælger så en anden adressat for sit tema. Der opretholdes en kooperation i samtalen mellem M og R, der kun fungerer på overfladen og kun for det tema, som M også accepterer. Den overfladiske kooperation opretholdes ved hjælp af formel imødekommenhed og høflighed fra begge sider.

Konversationel kooperation (fastslår Grice) er altid kooperation i forhold til et mål for samtalen: 'the accepted purpose or direction of talk exchange in which you are engaged' (Grice 1975:45). Det implicerer, at deltagerne i det mindste skal have en fælles opfattelse af, hvad der forhandles, og en fælles forståelse af situationen. Hvor – som i dette tilfælde – det potentielle resultat ikke er givet på forhånd, skal det forhandles lokalt under samtalen. Problemet er her, at kooperationen kun fungerer på overfladen og for det ene samtaletemas vedkommende, fordi det viser sig, at samtaledeltagerne har uforenelige målsætninger og forestillinger om, hvad der kan forhandles om: Rs målsætning er at holde samtalen i gang og undgå, at den bliver afsluttet, før han har fået sit budskab igennem. Ms mål er at undgå at blive involveret eller gjort ansvarlig for nogetsomhelt andet end turistinformation. Men samtidig er M professionelt forpligtet til venlig og høflig kundebetjening. At kommunikationspartnerne forfølger forskellige mål fører til, at M indskrænker sin kooperation til ren 'proforma-kooperation'. Sprogligt bestræber begge samtalepartnere sig imidlertid på at udsende positive signaler, der forhindrer et synligt brud, men tværtimod camouflerer et kooperationsdeficit under en overflade af velvilje og høflighed. De afslørende brudflader i samtalen er dels Ms rent fatiske *back-channel*-signaler, dels kampen mellem M og R om at lukke eller videreføre samtalen, dels Rs inddragelse af tilhørerne. På det rent turistiske plan fungerer kooperationen imidlertid, men kun for så vidt at Rs spørgsmål og anmodninger holder sig til det monotematiske plan, som M har afstikker.

Den høflige imødekommenhed har en beskyttende virkning for medarbejderen. Han undgår at ytre sig til det tema, han ikke ønsker at diskutere, samtidig med at hans professionelle image opretholdes. Man kan se denne samtale som en kamp om en 'dagsorden', om retten til at vælge hhv. fravælge et tema. I virkeligheden er det også en kamp om retten til at definere turistkontorets *activity framework*, og Rs forsøg på at bryde en institutionel defineret ramme for, hvad der kan forhandles, har ikke store chancer for at lykkes.

Hvis jeg skal drage interkulturelle konklusioner, må de gå i retning af, at der er tale om en kommunikationssituation præget af mangel på overensstemmelse i den implicitte sociale definition af mødet. Der er tale om parter fra samme kultur, men fra forskellige subkulturer, der har hver sin implicitte definition og forståelse af den kommunikationssituation, de befinder sig i.

*Syddansk Universitet,*

*Grundtvigs Allé 150, DK-6400 Sønderborg*

## Note 1

Jeg blev under diskussionen af mit foredrag på konferencen opfordret til at gå tilbage i samtalen og lede efter spor, der peger frem mod den efterfølgende konflikt (hvis begyndelse jeg lokaliserer til pausen i linje 29). Jeg mener ikke, at de er til stede. Som deltager i samtalen må jeg konstatere, at ingen af de tilstedeværende f.eks. har hæftet sig ved noget usædvanligt i forbindelse med Rs selvpræsentation (som formand for en tysk-jødisk forening på besøg hos foreningens æresmedlem i F.), men har opfattet den som et led i hans forklaring på, hvorfor han opsøger turistkontoret i F. De tilstedeværende synes næppe at have perciperet eller registreret den, hvad følgende båndudskrift af den efterfølgende samtale på turistkontoretmellem T, M og EF tyder på:

T: wer war das?
M: da fehlt eine stadtführung,
T: wer war das?
M: neh, ich hatte zwischenzeitlich [ein gespräch am telefon].
T: [Nein, nein. der der oberrabbiner,] wer war das denn?
M: von der jüdischen gemeinde, irgendwo?
EF:in Holland irgendwo.
M: ja genau! und er war hier zu besuch bei unserem Hermann Hansen. der ist heute 90 jahre alt geworden. da war die ganze stadtelite mit bürgervorsteher, und er kam jetzt natürlich gerade her, um sich zu informieren, neh?

Noget andet er, at selvpræsentationen rejser en sproglig ambiguitet i forbindelse med det 'wir', R anvender i linje 30: Den pronominale reference kan gå både på den forening, han repæsenterer, på ham selv og den ledsagende kvinde eller på begge dele.

## Litteratur

Brown, Penelope & Levinson, Stephen C. 1987: Politeness. Some universals in language usage. Cambridge etc.: Cambridge University Press.

Drew, Paul & Heritage, John (red.) 1992: Talk at work. Interaction in institutional settings. Cambridge etc.: Cambridge University Press.

Fredsted, Elin. 1998: Analyser af dansk og tysk talesprog. Oslo: Novus Forlag

Fredsted, Elin & Pedersen, Karen Margrethe. 1994a: Turistservice på tysk i Danmark. Flensborg: Institut für Regionale Forschung und Information (LINGUA-projekt): (a) undervisningsmappe (214 s.), (b) 2 lydkassetter (2 ( 70 min.), (c) diskette med ordlister, (d) undervisningsvideo (30 min) i tysk som fremmedsprog.

Fredsted, Elin & Pedersen, Karen Margrethe. 1994b: Turistservice på dansk i Slesvig-Holsten. Aabenraa: Institut for grœnseregionsforskning (LINGUA-projekt): (a) undervisningsmappe (192 s.), (b) 4 lydkassetter, (c) diskette med ordlister, (d) undervisningsfilm i dansk som fremmedsprog.

Goffman, Erving. 1974: Frame analysis. New York: Harper & Row.

Goffman, Erving. 1981: Forms of talk. Oxford: Basil Blackwell.

Grice, H. Paul .1975: Logic and Conversation, i: Peter Cole & Jerry L. Morgan (red.) (1975): Syntax and semantic. Vol. 3: speech acts. New York etc. Academic Press, s. 41-58.

Gumperz, John J. 1992: Contextualization revisited, i: Peter Auer & Aldo di Luzio (red.) (1992): *The contextualization of language*. Amsterdam, Philadelphia: Benjamins, s. 39-53.

Heritage, John & David Greatbatch. 1991: On the institutional character of institutional talk: the case of news interviews, i: Deidre Boden & Don H. Zimmermann (red.) (1991): Talk and social structure. Cambridge: Polity Press, s. 93-137.

Levinson, Stephen C. 1992: Activity types and language, i: Paul Drew & John Heritage (red.) (1992): Talk at work. Interaction in institutional settings, Cambridge etc.: Cambridge University Press, s. 66-100.

Psathas, George. 1995: Conversation analysis. The study of talk-in-interaction. London etc.: Sage.

Schegloff, Emanuel A. & Sacks, Harvey. 1973: Opening up closings, i: Semiotica 7, s. 289-327.

Wittgenstein, Ludwig.1953: Philosophische Untersuchungen. Frankfurt/Main: Suhrkamp.

# Overlap Management and Interactional Competence

*Karsten Gramkow*

## Abstract

In various language research disciplines there is a long-standing tradition for looking at conversational overlap as being violative of turn-taking norms. Until now, the general view of such simultaneous talk has been that it is basically competitive and interruptive. This paper discusses contributions from various researchers who have initiated new ways of looking at overlapping talk, and through analyses of a number of data extracts it is shown that such talk is generally supportive and collaborative. The paper claims that there is little difference between talk in a lingua franca and in a mother tongue, and that overlap is one of the conversation loci in which understanding and relationship-building are jointly constructed, in part due to the interactional competence of the participants.

## Communication and miscommunication

The data used for the analyses of this paper are drawn mainly from tape recordings of international business phone calls between non-native speakers of English. While such lingua franca situations seem to carry *some* real-world importance for the participants, in the sense that they, in their capacity as informants, feel inclined to indicate that by speaking English, they feel on 'common ground', I have found through my micro-analyses of their conversations that such non-native talk differs *only marginally*, if *at all*, from mother tongue interaction; overall, the conversational techniques used seem to be similar.

From a 'garfinkelian' point of view (Garfinkel & Wieder, 1992), it may be argued that international communication constitutes a 'perspicuous setting' for the study of cooperation, and indeed, the often-*presumed* problematic nature of such non-native speaker conversations has been the subject of many a sociolinguistic inquiry and analysis. Nevertheless, the conversations show few examples of communication which is ultimately unsuccessful; even in cases where the fluency of one or both speakers seems limited, they manage to get their business done, in a manner which they themselves display to be normal and meaningful.

A priori it has often been assumed that non-native speech-in-interaction makes for situations of misunderstanding and miscommunication. The same goes for simultaneous, or overlapping speech: interruptions are seen as violations of the standard turn taking pattern of one speaker-at-a-time, and simultaneity of talk as a *reason* for misunderstandings.

According to this tradition it might be argued that if such two socalled 'problem-filled' instances, i.e., non-native talk *and* overlap, coincide, even more problems would be expected to occur. This rarely seems to be the case in reality, however. Perhaps this is not all that surprising in instances of e.g. greetings or leave-takings, where the participants often, because of the *semi-ritualized* character of such interaction sequences, may be familiar with the mechanics of what happens, - but what with sequences of talk of a less prepatterned, more 'open-ended' character, as for instance talk including overlapping continuers or repetitions?

## Overlapping continuers

Through micro-analyses of a selection of short examples this section demonstrates how non-native speakers, ranging from *barely* to *very* fluent are able to *understand*, *issue*, and *use as resources* various conversational contributions, even in cases where these are uttered in overlap.

According to Schegloff (1982: 81), vocalizations like ‘uh huh’ exhibit acceptance of an extended unit of talk under way, and, as stated by Goodwin (1986: 213ff), at the same time such vocalizations deal with how units of talk are sequenced in relation to other units of talk: by issuing a continuer, the recipient *passes* the opportunity to take the turn; in this way he is *collaborating* in the achievement of speaker’s ongoing turn.

Continuers are often uttered in connection with turn constructional unit (TCU) boundaries, or at least in connection with what *the recipient* perceives as TCU endings:

Ex. 1: IK/HE/HE-MF/DK-Spain/1994/KGA/195-198

1 H: I'm chairing this: er:: Tempus session and that's all what I'm doing [::h ] they- they:
2 M: [uhhuh]
3 H: they all had a:m- a meeting they had a (.) th- the executive committee meeting in :h

M’s ‘uhhuh’ in l. 2 functions as what Goodwin (1986: 207) has called a ‘bridge’ between units. It is uttered close to the ending of a TCU; the original speaker, H, by performing an inbreath after ‘doing’, holds back the production of the next unit until M has finished the continuer, after which H is free to proceed. M in no way attempts a turn take-over; her contribution, while displaying attention to the ongoing action, also signals that she has received one unit and is waiting for the next to come. By continuing, speaker demonstrates that the continuer has been heard and is acted upon.

As exemplified in extract 2, speaker sometimes, though not always, breaks up or holds back his next unit until the continuer is finished, before he resumes the turn:

Ex. 2: Dantel/1/A2/VJ-T/DK-CH/1995-96/KGA/17-20

1 T: great (.) great=
2 S: =and we’ve ju-just got er: very little snow

3 T: [uhhuh]
4 S: [so: th-] :h the ground is white und er::

T's continuer in l. 3 occurs after the ending of S' TCU in l. 2; there is a short 'between TCUs-pause' between the end of l. 2 and the beginning of l. 4, before the next turn unit is uttered, and as S chooses to continue the turn at the moment when T utters his continuer, overlap occurs. But S, after uttering the first word ('so:') of the next unit, breaks off his utterance in mid-word on a 'th-' and holds back the rest until T's 'uhhuh' has been completed.

The participants display meticulous orientation to each other's talk, - they are able to monitor the talk of their coparticipant and coordinate their own contributions with his in mid-word. Even though their linguistic proficiency may be less than perfect, their *'interac tional performance'* remains intact.

In the next short sequence, P utters three 'yeah's', all of which occur in overlap:

Ex. 3: Seafood1/PH/3/DK-Spain/PH-F/1993/KGA/23-28

1 F: Fernando and I'm workin' o:n your ideas (0.4) regar[ding] if-er[: Centra] Vivo an' [also ]
2 P: [yeah] [ okay: ] [yeah]
3 F: the >hypermarkets but [let-< (.) gi]ve me some time.
4 P: [ y- yeah ]
5 P: ye:s >I know Francisco I know but your-< (.) I-h: (0.3) I- >first of all I tho- I thought
6 you-you were having holiday right now

While neither of the 'yeah's' meet what might be termed Goodwin's 'standard' require ments for continuers, in the sense that they do not occur at TCU-boundaries, they seem at least to be oriented towards and issued in relation to either *pauses* (the 0.4 second pause in the first case in l. 2), *a list of names* (the second 'yeah' of l. 2), or, in the last case (l. 4), to *a perceived TCU- ending* after 'hypermarkets'.

After P's two 'yeah's' and his assessment 'okay' in l. 2, F in l. 3 chooses to continue at high speed, but P has already set his acknowledgement token 'yeah' in motion. Realizing that his coordination was not perfect, P stutters ('y- yeah'), and F, upon hearing the overlap, pauses after 'let-' for a fraction of a second, before proceeding in the knowledge that P simply produced another continuer.

However, the added impact of the continuers in this case is a display of impatience, and in consequence, P in l. 5 starts his turn with a 'yes', followed by some topical, agenda- related work, for example by saying 'first of all', thus changing the subject slightly (l. 5 - 6). F does not object to P's taking over of the floor, and the conversation continues.

Extract 4 shows how continuers may be performed in series:

Ex. 4: Dantel/1/A2/VJ-T/DK-CH/1995-96/KGA/108-119

```
1    S:   you know he just asked [me ] (.) ::h if you had any :h remarks to the unit [:h  ]
2    T                           [hm]                                                [yea]
3    S:   because then it's time now (0.5) :h to ask for these changes if there are [some]
4    T: •                                                                            [yeah]
5    S:   changes you want [to  ] have
6    T:                    [yea]
7    T:   yeah (.) yeah >yeah yeah [yeah yeah<]
8    S:                            h before  ] he finalize the lay-out and he's just working on
9         the lay-out today.
10          (0.5)
11   T:   uhhuh
12   S:   ::h so if you can make your sketch er: he can take care er immediately
```

T's first 'yea' of the sequence, in l. 2, follows the standard pattern, with the continuer bridging two units. In contrast to this, the 'yeah's' in l. 4 and 6 both occur *in overlap with* an ongoing unit. The interventive use of the word 'yeah', together with the many repetitions of it in l. 7 may be a sign of impatience, of waiting eagerly for the sought-for informa- tion to be produced. If 'yeah' was to be viewed as a sign of *imminent speakership*, T would have taken the opportunity in l. 7, where S seems to have relinquished the floor. However, T does not start a turn here, and S proceeds, overlapping T's last two 'yeah's', in l. 8.

Another possibility is that T utters his series of continuers in order to win time to think about the offer put forward by S in the previous lines. At any rate, T seems satisfied with S' expansion and after a pause of 0.5 seconds in l. 10, he acknowledges S' contribution with the 'uhhuh' in l. 11.

## Repetition

'Yeah' is not the only conversational item repeated in these conversations. Overall, the repetition of words, sequences of words, or single syllables is used resourcefully, as a means to 'outwait' the coparticipant, as an element of a particular style, to ensure understanding in cases where it is considered to be threatened, as a joking element, etc. Often such repetitions occur in overlap, at least partly, as can be seen in the following extracts.

Repetition of words or parts of words may be performed by the same speaker (see above), but sometimes close monitoring of the other party's talk results in exchange fragments where conversation items of varying length are uttered by *both* speakers, often latched on to each other. Occasionally, this 'ping - pong' exchange has a joking effect:

Ex. 5: Seafood1/PH/2/DK-Spain/PH-M/1993/KGA/227-237

1 P: but the-the last two: (1.0) is forty seven
2 M: eh: hh:: just a second I'm going to che:ck it dadadada (.) :h
3 P: y-you have i- the direct number er er to to [er you ] can phone me directly
4 M: [Marina]
5 M: Marina Marina Marina [Mar-
6 P: [Marina [Marina Marin-]
7 M: [ah yeayeayea ]yeah [:h ] er 4 5 9 [9 3 6 3 ]0
8 P: [ye-] [9 3 6 3s]
9 M: >fortyseven<

10 (0.5)
11 P: f-fortyseven

M's 'dadadada' in l. 2 fills out the second of checking time, foreseen by her 'just a second' uttered earlier in the turn, but as no obvious result of the information search is produced, P continues, and in l. 3, by mentioning the existence of a direct number, offers a clue to M's search. However, P hesitates ('er er to to') in the middle of the line, and M steps in with help, namely the company name P was apparently looking for. As this is P's own company, they both know that P is aware of this already, but in this concrete situation M shows *support* and *close attention* by issuing the name. Taking up the slightly joking 'dadadada'- mode from above, M does the same with the company name in l. 5, while she seems to be searching for the phone number. The mode is echoed and continued by P in l. 6, where he takes over M's repetition series with the short overlap of 'Mar'.

M regains the turn with a new series, 'yeayeayeayeah' in l. 7. In l. 8, P with his 'ye-' attempts to follow the ping-pong pattern now established, but discontinues as M in l. 7 follows up with the number she has been looking for. P acknowledges this immediately by 'shadowing' in unison parts of M's list, once he has recognized the number: '9 3 6 3'.

In l. 11, P finishes off the sequence through a last repetition as an acknowledgement of the last part of the number ('fortyseven'), read aloud by M in l. 9, thus returning to P's initial indication of the number in l. 1, as well as to the standard list-calling procedure of acknowledging the list item by item (Jefferson, 1990).

The intricate pattern of the various types of repetitions and echos in this sequence demonstrates how finely attuned the participants are to each other's talk, and how overlaps of various types may be employed with different effects.

The mode and mood is constantly changing, without any metacommunicative comments from the involved parties.

The 'ping-pong' character may be even more pronounced than in the extracts above:

Ex. 6: Seafood/JEK-M/1993/JEK2/B1/DK-Spain/KGA/148-151

1 J: yes. (.) well- (.) I can send to you now that we'll be sending it next Wednesday
2 [(*) next (*)]
3 M: [next-e next] i:s (.) tralala fourth of Agust.
4 J: that's right.

As part of finding the date of next Wednesday, J and M show very close attention to the other's talk. J emphasizes her *next* in l. 1, and M in l. 3 overlaps with a repetition of the word, indicating a problem with regards to the date. J repeats the 'next' in l. 2 in overlap with M, who, after a short hesitation, re-repeats the word while she hears J start *her* repetition - see l. 2 and 3). After a bit of 'thinking time' in the form of ‘tralala’ (compare the 'dadadada' of the example above), the date is suggested by M and accepted by J in l. 4.

Repetition in the case at hand displays the participants’ ability to share communicative resources; the fact that it is used in overlap is not treated as problematic. Based on the overall smoothness of the exchange and the straightforwardness of J’s acceptance of M’s final suggestion it might be argued that the participants in this passage not only solve a conversation problem, but also ensure each other of their attention and mutual good will.

The repetition technique is also employed by less proficient interlocutors. In the next example, taken from a conversation about student accommodation problems between two university employees who did not know each other before the conversation, a few of the mispronounced words have been transcribed more or less as they are said, in order to highlight what would often, by analysts, be considered *potential* understanding problems.

Ex. 7: IK/MS/4/DK-Bel/1994/KGA/79-89

1 L: for the accommodation. h:[:h ] then er:m (.) she will receif (0.6) er: she will (.) have
2 D: [yes]
3 L: (.):h an: er- a groom (.) but not (.) er:: on the campice.
4 D: o:h no
5 L: cer[tainement]
6 D: [ okay ]
7 L: it [is a shange]
8 D: [ i-in a pri- ] pri-private house then
9 L: yes
10 D: yes :h
11 L: then she:: won't be with er: :h er: the friend.

L, who is Belgian, speaks with a very distinct accent (he uses 'receif' for 'receive', 'groom' for 'room', 'campice' for 'campus', 'shange' for 'change', and he code switches (l. 5: 'certainement'). This type of language proficiency problem is never made an issue by D. However, D *does* check her own interpretation of L's l. 3, starting in overlap in l. 8 by telling L what she thinks must be the natural consequence of the fact that the student will not be able to stay on the campus; namely that the student will be living in a private house instead. While the simultaneous speech here (l. 7 and 8) may be interpreted as a simultaneous start after D's 'okay' in l. 6, D demonstrates her sensitivity to the fact that she is speaking in overlap; she stutters slightly and repeats the syllable 'pri-' three times during and after L's utterance, thus outwaiting L before taking over the floor. L in l. 9 confirms D's interpretation of the situation, and in l. 11 offers an elaboration of the consequences for the student, thereby showing that D's overlap was relevant and has been accepted.

### Longer overlaps

Overlaps are sometimes considerably longer than has been the case in the examples above. This does not necessarily mean that they are interpreted as offensive, and floor take-over may be obtained via the same techniques as was the case in shorter overlaps:

Ex. 8: Seafood1/PH/3/DK-Spain/PH-F/1993/KGA/335-342

1 P: :h and he was last year too I think or was it- (.) yeah I think was last year too ::h so
2 he's- he's quite-e:rm he's a personality (.) :h I think in in in Spain>=but I don't
3 know< if if the effects would be that great also outside Cataluña it was just- (.) ::h
4 some[thing I was thi- (0.3) if we could ge- get him as a spons-]
5 F: [well I mean: (.) Bar- (0.4) Barbar ]lona (0.3) Barcelona?
6 P: mmhm
7 F: plays (.) the Spanish (.) league
8 P: mmhm

Such cases of longish overlaps are not uncommon in the corpus. However, they seem to be confined to those dyads where the participants know each other well, as is the case in this extract. F's intervention here (l. 5) is not completely unexpected, as P's previous talk in l. 1 - 3 has displayed a number of uncertainties, false starts, repetitions, etc. That opens up the turn space for F, although P does continue for some time after F has started his contribution. Eventually, P relinquishes his turn in the middle of the word 'spons-' in l. 4, and soon, through 'mmhm's' in l. 6 and 8, P acknowledges F's interventive action. Again there is an element of the outwaiting technique involved; F builds up to saying 'Barcelona', uttered correctly and in full *only after* P has stopped talking. This indicates that his production of the word is slightly impeded by the overlap, but after his request concerning P's understanding, indicated via the rise in intonation at the end of l. 5, he continues his turn.

## Repetition, style, and performance

Neil (1996: 127) calls conversations with for example such a frequent use of repetition a 'collaborative style' of talk, while Tannen (1984) prefers 'high involvement style'. Common to these two interpretations of the phenomenon is that repetition signals attention and close monitoring of the ongoing talk. Neil studied a group of low-proficiency speakers of lingua franca English and found that the use of repetition in non-native speaker interaction was in no way treated as a sign of dysfunction, as claimed by e.g. Tarone and Yule (1989: 107). As all the ethnic groups represented in her data corpus use repetition in a similar way, Neil concludes that repetition seems to be a universal rhetoric device.

The analyses of this paper confirm the claim made by Neil; repetition is widely used throughout the data corpus. It seems that at the lower end of the proficiency continuum, speakers use it mainly as a resource for the interaction itself, as well as for propositional content, for example by repeating parts of other-produced utterances as a means of displaying understanding. Further up on the continuum, there are several examples of the use of repetition in conversational 'ping-pong', or as part of joking behaviour. This requires a finely tuned collaboration between the participants. Furthermore, the employment of a strategy like repetition may in some cases point to the parts of the conversation where misunderstandings *might* have occurred, in the sense that the participants are doing whatever work they think is necessary to overcome such difficulties.

## Overlapping speech: some concluding remarks

A number of information-carrying functions may happen in overlap: continuers, assessments, questions, answers, repetitions, completions, etc. These phenomena are almost always treated by the participants as unproblematic and supportive, rather than as interruptive and competitive. The participants demonstrate their ability to decode, understand and act upon the message, and how they, within a split second, are capable of changing the course of their ongoing turn as a result of such interventions.

Overlapping speech is part of normal conversation; it is generally supportive, informative, and always coordinative and collaborative. Overall, there is a primacy of text rather than of speaker; the conversation as a whole, and often each turn, is produced via the joint, coordinated efforts of the participants, rather than as a series of single speaker utterances. There seem to be no precategorized limits to what may happen in overlap, but, as the examples have demonstrated, most types of overlap are structurally and contextually well defined, and they are made orderly and meaningful by the participants.

## The competence issue

In the light of the examples above, it now seems relevant to ask what kind of communicative competence is involved in the successful coordinative and cooperative work displayed here - is it what Canale and Swain (1980) termed the grammatical, the sociolinguistic, or the strategic competence, or, in other words, is it related to 1. the rules of grammar, 2. the rules of use, or 3. the knowledge of how to handle shortcomings in the grammatical and sociolinguistic competences?

As we have seen, even in cases where people speak a language other than their mother tongue, they are able to time and coordinate their own conversational efforts with those of their coparticipant. As lingua franca talk is little different from (and works as finely tuned as) mother tongue talk, it must be because the interactants, *despite* their potential lack of proficiency in (in this case) English, *in the situation* are able to apply their interactional resources meaningfully and adequately. At this point it should be remembered that non-native speakers are normally competent members in their own speech community/culture/ language. The activities performed in non-native interaction will therefore often be complex, but occasionally the talk may (at the same time) be characterized by incomplete grammar and vocabulary, etc. (cf. e.g. Mazeland, 1994).

At the heart of the framework of conversation lies its sequential organization, and based on the *primary context* established through this sequential organization, the partici pants understand and make use of what goes on in the talk. It follows that understanding is achieved on an *interactional* basis rather than as a result of the skills of the individual interactant only. If the participants work together in the building of a joint floor for conversation, it seems more fruitful to look at competence also from a *'joint effort angle'*, and perhaps speak of the *'inter-actional competence'* of the participants.

As the features of conversational interaction are viewed by the participants as informative and orderly, and as these features function as the basis of their inferences and as their context for further actions, the participants' ability to jointly recognize, produce, and make use of these features constitutes an important part of the participants' interactional competence.

The term *interactional competence* has been chosen here to set the type of competence in question apart from previous competence descriptions by e.g. Chomsky (1965, who employed a distinction between competence and performance), Hymes (1972, for whom communicative competence consisted of grammatical, psycholinguistic, sociocultural, and probabilistic competence), Canale and Swain (1980, mentioned earlier), as well as some of the ideas about cultural and intercultural competence put forward by e.g. Hannerz (1992), although of course it builds on these.

*Interactional competence* is not *only* the ability to express oneself in a way which is linguistically and socioculturally understandable to the other participant or participants in a conversation; it *also* includes the ability to draw upon coordinative and collaborational skills that may be learned through mother tongue interaction but seem to be applicable to lingua franca situations as well.

The idea put forward by this paper, then, is that in non-native conversations and mother tongue interaction alike, participants are able to draw upon this general, or generic, interactional competence, to overcome potential problems of ambiguity and misunderstanding, without rendering them interruptive, unusual or exceptional in any way.

*Hassel Allé 4, DK - 9000 Aalborg, Danmark*
*e-mail: gramkow@sprog.auc.dk*

## REFERENCES:

Canale, M. and Swain, M. 1980. Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. Applied Linguistics 1(1), 1-47.

Chomsky, N. 1965. Aspects of the Theory of Syntax. Cambridge, Mass.: MIT Press.

Garfinkel, H. and Wieder, D.L. 1992. Two incommensurable, asymmetrically alternate technologies of social analysis. In: Watson, G. and Seiler, R. M. (eds.). Text in Context. Contributions to Ethnomethodology. Newbury Park: Sage.

Goodwin, C. 1986. Between and within: Alternative treatments of continuers and assessments. Human Studies 9, 205-217.

Hannerz, U. 1992. Cultural Complexity: Studies in the Social Organization of Meaning. New York: Columbia University Press.

Hymes, D. 1972.: On communicative competence. In: Pride, J.B. and Holmes, J. (eds.). Sociolinguistics. Penguin.

Jefferson, G. 1990. List-construction as a task and resource. In: Psathas, G. (ed.).Interaction Competence. Washington, D.C.: University Press of America.

Mazeland, H. 1994. What Kind of Membership Knowledge is used in the Analysis of Foreign Language Data? Paper presented at the 'Conversation analysis of foreign language data' research symposium, held at Odense University, Denmark, June 1994.

Neil, D.M. 1996. Collaboration in Intercultural Discourse: Examples from a multicultural Australian workplace. Frankfurt am Main: Peter Lang.

Schegloff, E.A. 1982. Discourse as an Interactional Achievement: Some uses of 'uh huh' and other things that come between sentences. In: Tannen, D. (ed.). Analyzing Discourse: Text and Talk. Georgetown Univ. Round Table on Languages and Linguistics 1981. Washington, D.C.: Georgetown University Press.

Tannen, D. 1984. Conversational Style: Analyzing Talk Among Friends. Norwood, N.J.: Ablex.

Tarone, E. and Yule, G. 1989. Focus on the Language Learner. Oxford: Oxford University Press.

# Stereotypes in action

*Annette Grindsted*

## Abstract

Most research on stereotypes has been done within the domain of sociopsychology. I take the view that language is the dominant means by which they are defined, assessed and communicated. I also assume as stated by Kim and Gudykunst (1984) that 'you cannot not stereotype'. This, in turn, implies that language provides the key tool for communicating stereotypes cross-culturally and my specific aim is to describe what interactants representing one culture seem to do with their stereotypes in interaction with in-groups as well as with out-groups. Although the stereotyping process can be broken down into a number of specific cognitive activities, my primary aim is to analyse the process of making reflexions on interactive responses to stereotypic beliefs. The analysis of such interactive responses to stereotypic beliefs is not based on intercultural interactions, but on semi-structured interviews between in-groups on out-groups.

## Data

The interviews were made in Danish-owned companies in Venezuela with both Venezuelan and Danish staff. The questions concern inter-group perceptions on work-related issues, centering around goal-oriented activities in different departments and at different levels in the organization. Representatives from each of the two national groups were interviewed by country-fellow women in order to ensure as high a degree of reliabilty as possible. The Venezolanean interviews were made by Adriana Bolívar, Universidad Central de Venezuela, and the Danish ones by myself. Depending on the informants' readiness or ability to respond to the questions the interviews last from 20 minutes to 1 hour. We made 7 Venezuelean and 8 Danish interviews. The present paper is based exlusively on the Danish interviews.

## The stereotyping process

What is a stereotype? I take it to be the verbal expression reflecting any cognitive representation of a social group, in this case national groups or single persons as members of these groups 'Kim and Gudykunst (1984)', 'Mass and Arcutri (1996)'.

To put words to a stereotype implies that the individual has passed through cognitive processes covering a wide range of activities, including how he perceives the world around him, how he classifies is, and how he gives it meaning. This is what Kim and Gudykunst (1984) refer to as psychocultural processes. It is widely accepted that these processes include 'Fant and Grindsted (1995):

*Generalized perceptions of the other party's behaviour*: When people meet people who belong to another cultural group, they are likely to use language to organize or systematize their perceptions in order to create representations of the other party that serve the purpose of predicting the other party's behaviour.

*Valuation of the other party:* A normal step in stereotype formation is through language to attribute a positive or negative value to the observed behaviour and to underlying values. This process necessarily involves matching the value system of one's own membership group with that of the other party's membership group.

*Inferences concerning underlying factors:* As rational beings it is natural to make attributions about factors thay may determine the observed behaviour.

*Defense to protect stereotypes against disconfirmation:* When you try to defend existing stereotypes against disconfirmation the motive is to ascertain that your perceptions and valuations are not only your own but are shared by at least some members of your own cultural group. In this way you achieve in-group back-up.

These steps reflect an idealized representation, and seems to indicate that the stereotyping process is a personal experience, which is of course, not the case. Stereotypes are often inherited or socially transmitted from members of your own cultural group. However, inherited, transmitted or not, people cannot get rid of them. They have to interact with them, deeply rooted in their minds. Since it is assumed that it is almost impossible to modify stereotype content, we are back at the beginning where I raised the question:What do you do with your stereotypes in interaction? Consequently, it seems reasonable to suggest that yet another process should be added to those already mentioned and that is '*reflexions on interactive responses to stereotypic beliefs*'. In other words the category addresses the question of what you do with your stereotypes in interactions with representatives from the other culture.

## Methodology

Bearing in mind the aim of the study I can find at least two different empirical approaches that I will briefly discuss. One is recordings of authentic interactions between representatives from different cultures. The other is interviews, in which the interviewer asks the interviewed about his personal experience. What can you do with each of these approaches?

## Authentic interactions vs interviews

Authentic implies that you can observe directly what happens between the parties interacting. There is no intervention from a third party. In interviews, on the other hand, the interviewer - through the questions he asks - is the origin of the messages given, and in this way he guides the informant through his own selective procedure. In this sense authentic interactions are non-selective, since it is impossible to tell in advance what kind of stereotypic behaviour can be observed in which kind of interactions. This implies that authenticity validates the results of the analyses. However, turning to the analysis of authentic material, at least two questions must be raised. One concerns the analist's possibilities to distinguish between what kind of communicative behaviour is generated through stereotyping processes and what is not. Earlier research has shown that certain types of intergroup situations seem to activate sociolinguistic stereotypes, such as mixed-sex interactions, caregiver-patient interaction, or high class speakers toward low-class speakers 'Giles and Coupland (1991)'. The important question to be answered is whether a discourse analytical approach to authentic international interactions will allow us to tell which communicative choices are driven by stereotypic beliefs and which are not. The second issue concerns analytic level. When analysing authentic interactions, you focus on what happens in one specific context, directing your attention towards the micro-level of discourse. This enables you to analyse the cognitive organization of interaction, but only certain aspects of identity-maintenance functions. In this way it has been shown that depending on the type of intergroup relation and the strength of cultural group identity, people will tend to shift their speech pattern either toward (i.e. convergence) or away from the person being addressed (i.e. divergence) 'Giles and Coupland (1991)'. But when you talk about the identity-expressive function of interaction, certain restrictions are imposed on you. One aspect is that you cannot analyse what people don't do (avoidance) or what they do later as a consequence of the interaction you observed. Another aspect is that you only to a limited extent can answer questions about why people behave the way they do.

Turning to the interview, I have already mentioned that the informant orients himself towards the selection procedure made by the inteviewer. In the same vein, concerning reliability, the interviwer - who is restricted from making direct field-observations - depends on the informants readiness to respond openly and objectively. On the other hand, the interview makes it possible to relate communicative behaviour with macro-contextual dimensions suchs as economic and political relations. Moreover, it allows you to make a more in-depth analysis of discourse management strategies and identity maintenance functions touching issues of inferences of why you think people behave as they do. One more aspect is that your informants can surpass the limits imposed by specific contexts, and in this way are able to inform on personal experience in general terms. A case in point is that informants can tell about communicative choices they do not make and the reasons why.

## Interviewing technique

The method chosen for this project is semi-structured interviews where two types of questions are asked. One type is about specific trait ascriptors related to interactive concepts regulating the interelationship between ego and alter. (Are the Venezolaneans polite, trustworthy, cooperative, assertive, etc.?). These questions are all closely related to the first - and only to the first - category of the stereotyping process mentioned before. The other type concerns the informants' perceptions of how specific types of interactive events are organized (meetings, negotiations, etc.).

The content of national stereotypes are categories that are known and widely accepted already, and one important argument that supports the idea of asking questions, where the answer is known beforehand is to neutralize to the widest possible extent the interviewer's personal influence on how the informant responds to the questions. The interviewer's role is of course to elicit as much information as possible, but not to elicit preferred answers. In this way the discourse regulating devices the interviewer has at his disposal should be reduced to the prepared questions (to elicit information), repair questions (if the question has not been properly understood or if he does not understand the reply), back-channeling devices (to confirm that he is an attentive listener), pauses (to elicit more information) asking next question because the interviewer finds that the previous question has been fully answered, or to put it in another way, that in the stream of talk produced by the respondent no new information is added.

In my analysis I have not taken into account how interviewer and respondent jointly construct the interaction. It could be argued however, that this type of interviews is an instrument that unconsciously is used to protect stereotypes against disconfirmation, and in this way serving the function of stereotype transmission and maintenance 'Briggs (1986)'.

## Respondents' setting up account slots

A minimal response to the interviewer's question concerns the explicitly elicited information and nothing more, i.e. confirming or disconfirming the validity of a trait ascriptor or ascribing traits to interactive events. These responses corresponds to the initial step of stereotype formation.

### *Generalized perceptions on the other party's behaviour*

*Example 1*

I opfatter du dem som individualister [øh] på arbejdspladsen

R nej det er de ikke [det gør du ikke] nej

I altså som modsætning til kollektivister ikke

R nej

*I do you think they are individualistic [eh] in the work-place*
*R no they are not [you think they are not] no*
*I as opposed to collectivists you know*
*R no*

In this example you the respondent *(R)* gives a minimal response disconfirming the trait ascriptor suggested by the interviewer *(I)*. In the first place *I* tries to elicit more information repeating *R's* utterance. And since *R* gives the same minimal response, *I* tries to reformulate the question contrastively. But again, *R* delivers a minimal disconfirming response. The first comment to be made is that *R* knows that he engages in the activity *interview* because *I* expects him to possess expert knowledge on the topics she introduces. Secondly, it is assumed that he knows that *I* is familiar with steretype type content. Consequently, he is supposed to be aware of the fact that in his response he adds no new information, as he is expected to do. In this way, to give only minimal responses to trait ascriptors must be considered a rather marked option. (I've only made a preliminary transcription and inform you of preliminary results. I've not transcribed pauses, which may be significant in this example, reinforcing the idea of markedness).

This means that the unmarked option is that the respondents add new information on their own initiative, which is, in fact, also what happens in most cases. This brings us to ask the question of how respondents construct account slots. The preliminary analysis of my interviews reveal that verbal expressions of the cognitive processes outlined above are all amply represented, as can be seen from the examples that follow:

### *Valuation of the other party*

*Example 2*

I øh ok vi kører videre øh hvad synes du så om deres sociale: deres sociale kompetence øh deres evne til at indgå i altså [sociale sammenhænge] ja

R der vil jeg mene der vil jeg mene den er absolut den er utrolig høj [den er høj (ler)] den er meget høj jeg mener at: jeg tror at der er meget få ting øh øh rent kulturelt der falder der falder nemmere for venezoelanere end det og at holde fest og komme hinanden ved [m'm] og tit så vil man som dansker nok synes at de overdriver en del [m'm] især når man tænker på effektivitet som trods alt også skal være til stede i en organisation ikk [ja] i på en arbejdsplads ikk

*I* *eh ok let's go on eh what then do you think about their social their social competence eh their abilitiy to enter into you know [social relationships] yes*

*R* *I think I think it is absolutely is is incredibly high [it's high] (laughing) it is very high I think that I beleive there are very few things that eh eh from a cultural point of view comes comes easier to venezolaneans than to have a party and care about each other [m'm] and often as a dane you will probably think that they are exaggerating quite a lot [m'm] especially when you think of efficiency, which after all also must be taken into consideration in an organizaion don't you think [yes] in the workplace don't you think*

The first part of the response it to add a trait ascriptor. Unprovoked R expands on it, making a negative assessment which is ethnocentric in character being based on what is meant by efficiency in a working context according to Danish standards. R marks completion by the tag question 'ikk'.

### *Inferences concerning underlying factors*

*Example 3*

I .....er det nemt at: at opbygge et venskab

R ja og nej altså: ja og nej jeg vil jeg vil godt lige sige i forhold til andre lande jeg har rejst i der er venezoelanerne sgu lidt mere: der er lidt af det her oliesheike over dem fra gammel tid hvor de var vant til det er så femten tyve år siden de havde de penge mellem hænderne men der sidder stadig væk noget af det der i dem at at: det var dem der var lisom lidt af sheikerne og og og andre mennesker de: de må bare indordne sig under forholdene ikk der 'har de altså lidt af det der i sig så 'stadig væk så jeg vil sige ja og nej alt efter hvem du kommer i kontakt efter hvor moderne de er indstillede sådan [hvad] i øverste etage som det hedder

*I* *......is it easy to make friends*

*R* *yes and no yes and no I would say that compared to other countries I have travelled in well the venezolaneans are more a little more like the oil sheiks from previous times where they were used to well it's fifteen or twenty years ago have a lot of money but something of that is still left of being that they were something like the sheiks and and then other people had to adapt themselves to the sitution you know there is still something of that left so I would say yes and no depending of whom you meet and how modern they are in the upper storey as it is called*

This example demonstrates a relatively high degree of cultural awareness on the part of the respondent, where he in his cultural attibutions relates situational factors to psychosociological factors in a historical perspective. Oil profits have created oil sheiks who have internalized a feeling of superiority in their interrelationship with other people. This attitude can still be found although the oil money is gone.

***Defense to protect stereotypes against disconfirmation:***

*Example 4*

I ok æhm synes du de: hænger sig i formaliteter og:

R ja utroligt meget [det gør de] utrolig meget

I hvordan: kommer det så til udtryk i det daglige

R ja altså især hvis de får en lille en lille smule øh: skal man sige en halv titel på ryggen eller et eller andet så: æh jeg har haft et par eksempler hernede med en *vi andre sad og grinede lidt af* men ja da han lige pludselig han rykkede en en en grad op og altså da så sku han jo ha sit eget kontor ikk *og gud døde mig om* han ikk var inde og måle op om det var fem centimeter mindre end naboens ikk *[ler] altså det er jo til at skrige af grin af ikk men altså*

*I ok eh do you think they are concerned with formalities and*
*R yes very much [they do] very much*
*I how is it reflected in daily life*
*R well yes especially if they get are promoted and they get something like half a title to show on their back you know I have seen a few examples here with a the rest of us were laughing a little at it but well he suddenly was promoted and he got his own office and good grief you know what he did he measured it to see if it was 5 centimeters smaller than that of his colleague [laughing] we couldn't help laughing but you know*

Talking about defense it is important to note that what is referred to is in-group back-up from peers at the work-place, and not about the relationship between interviwer-respondent. What he refers to is in-group behaviour that serves the function of stereotype maintenance about formality as a means to preserve hierarchy.

***Interactive responses to stereotypic beliefs***

*Example 5*

I øhm hvad med mødekultur de det er jo normalt har jeg næsten forstået at man [mødekultur møde til tiden kultur] m'm

R det kan jeg udtale mig som ekspert i jeg kommer altid for sent og det har jeg lært herovre ikk [begge ler] øh det generelle indtryk er at folk kommer: dryssende som de nogenlunde har lyst til min erfaring er ikke [mm] at de gør det min erfaring er snarere at vi gør det [m'm m'm] ja ja

*I eh what about meeting culture as I have understood it is quite normal [meeting culture to be on time] mm*

*R I can comment on that as an expert I am always late and I have learnt that here you know [both laughing] eh the general impression is that people saunter in more or less as they like my experience is that they do not [mm] my experience is rather that we do [m'm m'm] yes*

This example is almost an example on stereotype reversing in the sense that the hetero-stereotype is converted into an auto-stereotype.

I would like to clarify what is meant by this category. First of all, it does not relate to the interelationship between interviewer and respondent, but what the respondent tells the interviewer about his interactive responses to Venezolaneans. Secondly, it is important to bear in mind that we are talking about interactive responses. In the interviews, responses to stereotypic beliefs like 'this irritates me' or 'it is frustrating' are not interactive since they reflect momentary internal emotional states, that even if they can be immediately observed by the other party, are not intentionally transmitted to him. Let me give you a few examples of what is meant by this category:

| | Stereotypic beliefs | Interactive responses |
|---|---|---|
| 1) | Venezolaneans are unpunctual | I've learnt to be late |
| 2) | unstructured in conversation | make them stick to the point<br>present agenda |
| 3) | always simultaneous talk | cut through to make them speak one at a time |
| 4) | make exaggerated intitial positions in negotiations | I've learnt to do the same |
| 5) | nothing happens when you make an agreement | need to be whipped |

Earlier research on stereotype-based interaction has been done in the field of of the identity-expressive function of language in the framework of speech accomodation theory 'Giles and Coupland (1991)'. It has especially been directed towards shifts in speech patterns or speech styles, based on analysis of authentic interactions. However, the general assumption that depending on the type of intergroup relation and the strength of ethnolinguistic identity, people will tend to shift their speech pattern either towards or away from that of the person being adressed. Thus, both convergence and divergence are taken to be driven by sterotypic beliefs about how in-group and out-group members typically talk, and the direction of these speech shifts will depend on different factors such as striving for recipient's social approval or stressing one's own social identity.

In the present study interactive responses will be understood not only as reflecting the surface level of discourse but also underlying processes that determine interactive choices. When this is possible it is due to the empirical data collection method chosen, i.e. the interview, in which such issues can be explicitely touched upon. Another comment that needs to be made is that we are prevented from analysing the micro-level of discourse, since informants will be unable to answer issues pertaining to this level. So we are exclusively concerned with macro-organization. Following this line of thought, 2 such strategies have been introduced by the examples mentioned before. Thus examples 2), 3), and 5) reflects *assertiveness strategies*, where the informant states that he underscores or exaggerates perceived auto-stereotypical behaviour. Examples 1) and 4), on the other hand, reflects *accomodation strategies*. In this case the informant states that he adapts his behaviour to the perceived hetero-sterotype 'Fant and Grindsted (1995)'.

However, as was earlier mentioned, when comparing the analysis of authentic interactions with interviews, the latter method makes it possible to introduce yet another strategy, i.e. *avoidance*. This strategy, which for obvious reasons it is difficult to analyse in authentic interactions since it refers not to strategies that were chosen, but to what was left unchosen, refers to the inhibiting of responses because they are thought to be threatening to the other party. Let me give you an example:

*Interactive responses to stereotypic beliefs - Avoidance strategy*

*Example 6 (Kaj Simonsen)*

I .....hvis vi skal sammenligne med Spanien for eksempel der er det mit indtryk at når du kommer til spanien og kigger på interaktioner mellem spaniere så er der et meget højt element af selvhævdelse i det

R tværtimod øh en venezoelaner har et et et meget sjovt træk et: træk der nærmere ligner selvudslettelse ihvertfald når det drejer sig om den nationale karakter [ja] ja de kan: sågar have en ironisk til det at være venezoelaner hvilket igen set ud fra min min synspunkt er vældig charmerende [har øh] *man skal bare passe på ikke at misforstå det man skal ikke komme som [m] udlænding og kritisere den venezoelanske nation* det er de gode til selv [ja] men de kan godt grine af det

I if we compare with Spain for instance it is my impression that when you are in Spain and observe interactions between Spaniards there is an important element of assertiveness

R on the contrary a Venezuelan has a very funny trait a trait that is closer to self-destruction, at least when it is about the national character [yes] yes they can be ironic about being Venezuelan which in my point of view is rather charming [have eh] but you have to be careful and not misunderstand it as a foreigner you should not criticize the Venezuelan nation they are good at doing that themselvs [yes] and they can laugh at it

The avoidance strategy is that you should not criticize their self-destructive behaviour - and as I understand it what is meant is that in interaction you should neither confirm nor disconfirm the national image of self-destruction, but, for instance, simply change the subject.

## Conclusion

Let me sum up what I have been trying to demonstrate. It has been demonstrated that through a discourse analytical approach to the study of interviews on intergroup perceptions on work-related issues it has been possible to identify the psychocultural stereotyping procesess as suggested by e.g. Kim and Gudykunst (1984). The analysis seems to support the idea that that the process of reflexions of interactive responses to stereotypic beliefs should be added to the standard processes mentioned by Kim and Gudykunst. In this latter category, it has been found that the informants choose among three different strategies: assertiveness, accomodation and avoidance. These are the preliminary results in an on-going project. Further issue to be explored are the Venezuelans' stereotypes about the Danes, and to analyse, as far as it is possible, through direct observarion (video-recordings) the importance of stereotypes in authentic interaction.

*University of Southern Denmark*
*Main Campus: Odense University*

## Transcription conventions

| | |
|---|---|
| [...] | back-chanelling devices |
| : | continuative intonation pattern |
| ' | emphatic stress |

## References

Antaki, Charles. 1994. Explaining and Arguing. The Social Organization of Accounts. London: Sage Publications.

Briggs, Charles L. 1986. Learning how to ask. A sociolinguistic appraisal of the role of the interview in social science research. Cambridge University Press.

Fant, Lars M. and Annette Grindsted. 1995. Responses to value mismatch in intercultural negotiation interaction. Text 15(4). 561-588.

Giles, Howard and Nikolas Coupland. 1991. Language: Contexts and Consequences. Pacific Grove, California: Brooks/Cole Publishing Company.

Gudykunst, William B. and Young Yun Kim. 1984. Communicating with Strangers: An Approach to Intercultural Communication. New York: Random House.

Mass, Anne, and Luciano Arcuri. 1996. Language and Stereotyping. Ed. by C. Neil Macrae, Charles Stangor and Miles Hewstone. New York: The Guilford Press.

# Danish in Greenland - a new dialect?
## Preliminary results from a pilot project

*Birgitte Jacobsen*

**Abstract**
In a pilot project I want to investigate whether there is a specific variety of Greenland-Danish, spoken by the monolingual Danish-speaking teenagers who have grown up in Greenland. This variety might be considered a 'languages in contact' phenomenon, emerging in the contact between Danish and Greenlandic, supposedly via the more or less Greenlandic accented Danish of the bilinguals. The bilinguals have high prestige in many, though not all, respects, and this may possibly explain the emergence of new norms for the spoken language, different from the norms in Denmark. Presently I focus on phonological matters, but later other matters will be included.

## 1. Introduction

The linguistic constellation in Greenland, and in particular in Nuuk, is complex. So are the concepts of ethnicity and identity, which are not in any simple way related to language. Nuuk is the capital of Greenland, and as an administrational and educational center it has a great number of Danes; in Nuuk 33 % of the inhibitants are Danes, as compared to 14 % in the rest of Greenland. In 1953 Greenland changed from being a colony to becoming an equal part of Denmark, and during the sixties Danish became more and more widespread, sometimes at the expense of Greenlandic; actually there are quite a few young Greenlanders today who do not speak Greenlandic. During the seventies a reaction came, and Greenlandic gained ground, sometimes at the expense of Danish. To a certain degree Greenland is a diglossic society.

In the course of history Danish and Greenlandic took turns with respect to having highest status. Today Greenlandic has the highest status in certain respects, Danish in others; but bilingual persons tend to have high status in most respects, although not in all.

Danish is a minority language, but in a special kind of minority situation: Greenland is a society where the majority also to a great extent speaks the minority language; further, it is in many cases possible to live and fulfill your functions in the country not knowing one single word of the majority language.

## 2. Purpose, working hypothesis and design

In this project I will investigate whether there is a particular spoken variety of Danish among the monolingual young Danes who have grown up in Greenland. My project is similar to the investigations of spoken Finnish in Sweden (cf. Lainio 1989; Wande 1991); in some respects my data are parallel to the so-called Rinkeby-Swedish (cf. Kotsinas 1987), and to the so-called Kebab-Norwegian (cf. Aasheim 1997). - The question of a specific Greenlandic variety of Danish has not been investigated before, although it has been touched upon loosely in a discussion of 'semilingualism' (cf. Langgård 1992, Trondhjem 1998).

My working hypothesis is that there has emerged a new variety of Danish in Greenland via the Danish of the bilingual Greenlanders. It is 'dialects in contact' situation, indirectly a 'languages in contact' situation. The influence can be explained in terms of the bilinguals having higher status, causing the emergence of a new norm.

I compare the language of the speakers in Greenland with the language of speakers in Denmark to avoid dependence upon a rather vague and possibly outdated 'standard Danish' as a frame of reference. - My main focus is on phonetics and phonology, with special emphasis on prosodic features. The independent variables include, apart from those related to the individual, also conversation type and level of formality. The linguistic pattern should be analyzed both in terms of groups and in terms of individuals.

If the pilot investigation gives a positive answer and I do find a specific variety of Danish in Greenland, the next steps will be acoustic investigations on the one hand, and on the other a more detailed analysis of speaker variables, language use, code choice, network, and language attitudes. Also I will include lexical, syntactic, morphological and pragmatic matters.

## 3. Procedure

The pilot investigation includes questionnaires, class-room observations, and tape recordings. In the questionnaire the pupils were asked different person-related questions; further they were asked to rate their own linguistic competence in both Danish and Greenlandic, and to tell which language they speak with different familiy members and with friends at home and friends at school. Concerning linguistic competence the pupils were asked to rate their competence in speaking and in understanding Danish and Greenlandic respectively. They were also asked which language they judge themselves to speak at the highest level. This is not as redundant as it might seem, because even if a person rates his/her two languages equally high, s/he might feel more comfortable with one of the two.

On the basis of the questionnaires I selected the persons for taperecordings. This way of selection ensured that the sample was random, but stratified according to the relevant variables.

## 4. Subjects

In the pilot investigation I restrict myself to speakers from grade 9 and 10 and the first two classes in the gymnasium. Later I will include younger speakers as well. 169 pupils answered the questionnaires, and 33 pupils were selected for tape recordings. The recordings from Denmark comprise 28 pupils.

## 5. Data

The tape recordings consist of three different conversation types: 1) An individual interview of 5-10 minutes. 2) Reading aloud an easy piece of prose. 3) Group conversation between 3-4 persons, lasting 10-20 minutes.

The primary purpose of the prose reading was to get a basis for direct comparison; also, it would give a differentiation of formality, prose reading probably yielding the most formal style, group conversation the least formal style, and the interviews something in between. However, in the pilot investigation there are not enough data for a thorough style analysis.

## 6. Preliminary results

### a. Nuuk-Danish - what is it like?

Turning now to the first question: is it possible to distinguish a specific variety of Greenlandic Danish spoken by the young Danes in Greenland? - the answer is 'yes'. First, their Danish is different from the Danish spoken by Greenlandic bilinguals, who has a more or less marked Greenlandic accent. Second, this variety spoken by young Danes in Greenland is also different from the Danish spoken by young Danes in Denmark. - For reasons of convenience I will call the specific Danish of the Danes in Greenland 'Nuuk-Danish', and the Danish of the control group in Denmark I will call 'Denmark-Danish'.

Nuuk-Danish has a specific accent - but it is not immediately evident that it is a Greenlandic accent. The most salient feature is a specific staccato and monotonous way of speaking. - The following is an attempt to explain the prosodic pattern of Nuuk-Danish in terms of interference between two different prosodic systems:

It is tempting to label Nuuk-Danish as being 'syllable-timed' - as opposed to Denmark-Danish which could be labelled 'stress-timed'. However, Anders Eriksson (1991) has shown that stress-timing and syllable-timing are not a matter of isochrony; but he found that there *is* a perceptual and structural difference: the two language types in question differ mainly with respect to *relative difference* between stressed and non-stressed syllables, and with respect to *syllabic structure*. In the so-called 'syllable-timed' languages the difference between stressed and non-stressed syllables is relatively small, while the so-called 'stress-timed'languages have relatively more complex syllable structure and a relatively greater prominence of stressed syllables.

Turning to Danish, the stress group pattern is charaterized by a pitch movement from the stressed syllable to the posttonic unstressed ones, rising in some dialects, falling in others, roughly speaking (cf. Grønnum 1992 for a thorough analysis of Danish prosody). Stressed vowels have greater duration than unstressed ones, they may have 'stød', and they may have a different quality.

In Greenlandic there is no phonetic category of stress (cf. Rischel 1974: 91ff, Jacobsen forthcoming); a word is intonationally characterized by a terminal tonal contour with a salient pitch drop on one of the last two morae, depending on the word's position as either phrase final or phrase internal. (For more details cf. Mase 1973, Rischel 1974: 79ff, Fortescue 1984: 340ff, Nagano-Madsen 1993, Nagano-Madsen & Bredvad-Jensen 1995, Jacobsen forthc.)

Greenlandic has distinctive quantity in both vowels and consonants, so there are four different syllable types with increasing syllable weight: (C)V (short, open) (C)VC (short, closed) (C)VV (long, open) (C)VVC (long, closed). The four different syllable types differ primarily in the durational parameter. The syllabic structure is rather simple: there are no clusters of three consonants, and only one (surfacing) cluster of two consonants, so there are only simple codas (syllables are (almost) only closed by long consonants).

Prosodically Greenlandic is characterized by the rhythm which relates to syllable weight and to word final intonational contour. There may be a certain impression of stress on the mora bearing high pitch, but it is really a matter of tonal prominence, not stress proper. The lack of a stress distinction together with the simple syllabic structure means that Greenlandic does not sound like a 'stress-timed' language, but rather like some sort of a 'syllable-timed' language.

When a Greenlander, who lacks the distinction of stress in his own language, is supposed to express a stress-distinction in a Danish word, there seems to be two alternative strategies: one is a tonal exaggeration, where usually also other prosodic features like 'stød' are exaggerated. The other is a reduced pitch movement. One way of explaining the pitch reduction could be: because pitch movement in Greenlandic is a word or phrase boundary marker, a Greenlandic speaker might be reluctant to use pitch movements in the beginning of words, or rather, a Greenlandic speaker may 'suspend' the tonal movements until the end of the phrase. Consequently the pitch range throughout the phrase will be reduced, resulting in a monotonous impression. This serves as a 'link' to Nuuk-Danish.

Nuuk-Danish also has a specific staccato speech rhythm, so there is probably also a parameter of duration on stage. The interplay between pitch and duration will of course be instrumentally investigated later.

The specific prosody of Nuuk-Danish can thus be described as *a reduced acoustic difference between stressed and non-stressed syllables*. This might be explained in terms of interference between two different prosodic systems, perhaps *transfer from a 'syllable-timed' system to a 'stress-timed' one.*

### b. Nuuk-Danish - who are the speakers?

*Who speaks Nuuk-Danish? (NB: preliminary)*

| *Parents* | *Category of speaker* | | | | | |
|---|---|---|---|---|---|---|
| | D | Dg | **D**/G | D/G | D/**G** | dG |
| Grl/Dan | + | + | + | ? | - | - |
| Grl/Grl | - | - | - | - | - | - |
| Dan/Dan | +/- | + | + | - | - | - |

D: Danish speaking, monolingual; knows nothing or little Greenlandic; never uses Greenlandic
Dg: Danish speaking; knows a little Greenlandic; uses Greenlandic occasionally
**D**/G: Bilingual, rates Danish as best language
D/G: Bilingual, equally competent in both languages
D/**G**: Bilingual, rates Greenlandic as best language
dG: Greenlandic speaking; knows a little Danish; uses Danish occasionally

As far as I can see now, a typical Nuuk-Danish speaking person is a young person with Danish as his/her best language, s/he may or may not speak Greenlandic, and s/he has either a Greenlandic mother or a Greenlandic father. Another typical Nuuk-Danish speaking person is a young person with Danish as his/her best language, but s/he speaks Greenlandic - and s/he may or may not have a Greenlandic mother or father. However, also some monolingual Danes with both parents being Danish speak Nuuk-Danish. It can be seen that there are two important dimensions: an ethnic and a linguistic, not necessarily interdependant.

Let us for a moment look at those young Greenlanders who define themselves as primarily Greenlandic speaking; they typically have a specific Greenlandic accent with direct phonological interference (e.g. exaggerated retracted vowel quality in front of *r*). I know of more than one case where a young Greenlander who grew up as only Danish-speaking, and who struggles to learn to speak Greenlandic as a grown-up, suddenly starts to speak Greenlandic-accented Danish. It is tempting to interprete this accent as a way of signalling a Greenlandic identity.

Turning to the Nuuk-Danish speaking young Danes, one could interprete their accent along these lines and suggest that their way of speaking signals Danish ethnicity with a Greenlandic solidarity, i.e. signals that they are Danes, but belong to the country.

### c. Nuuk-Danish - when is it spoken and to whom?

Although style and conversational partner are not independent, one may see certain tendencies. For those speakers who speak both varieties there is a tendency to use Denmark-Danish rather than Nuuk-Danish in prose-reading, and there is a tendency to use Nuuk-Danish rather than Denmark-Danish in group conversations. These patterns are expectable: choice of code depends on level of formality.

But code choice of course also depends on conversational partner: speakers tend to use Nuuk-Danish to a greater extent when speaking with peers than when speaking with me. It is an obvious evidence of signalling solidarity with the group. It was especially evident in the case of an American boy, who was born and has grown up in Nuuk (both parents being American). His first language is American, but he speaks Danish fluently, and only a little Greenlandic. When he spoke to me, he spoke Danish with an American accent - but when he spoke to his peers, he spoke Nuuk-Danish!

## 7. Concluding remarks.

There seems to be a specific variety of Danish in Greenland - at least in Nuuk. Certain prosodic features of this variety can be interpreted as interference from the Greenlandic prosodic pattern.

The speakers of Nuuk-Danish usually have a certain relation to Greenland, be it ethnic and or linguistic. But some of the speakers have no such relations. Where Danish with Greenlandic accent signals Greenlandic ethnicity, Nuuk-Danish seems to signal Danish ethnicity with Greenlandic solidarity.

I have called my paper: Danish in Greenland - a new *dialect*? The term is chosen for reasons of convenience. It is not a dialect in the classical sense. It may rather be called a regional variety, "regionalsprog", in the Danish sense (cf. Kristiansen 1992, Petersen 1994). The emergence could be explained like this: Nuuk-Danish is a new urban variety, which has emerged as a reaction to the increasing critical attitude towards Danish dominance: in a period where it is important for many people to emphasize their Greenlandic ethnicity, the Danes with close relations to Greenland face a need to signal on the one hand that they are Danish, but on the other to emphasize that they belong to Greenland. In that sense Nuuk-Danish may be called an ethnolect (cf. Appel-Muysken 1987).

*Dept. of Greenlandic Language and Literature*
*University of Greenland*
*P.O. Box 279, DK-3900 Nuuk*
*Tlf: (+299) 32 45 66, Fax: (+299) 32 47 11*
*E-mail: bj.unigreen@greennet.gl*

## References

Appel, René & Pieter Muysken. 1987. Language contact and bilingualism. London: Edward Arnold.

Berge, Kjell Lars & Ulla-Britt Kotsinas (utg). 1991. Storstadsspråk och storstadskultur i Norden. Föredrag från ett forskarsymosium. MINS 34. Stockholm.

Eriksson, Anders. 1991. Aspects of Swedish Speech Rhythm. Göteborg, University of Göteborg.

Fortescue, Michael. 1984. West Greenlandic. London, Croom Helm.

Grønnum, Nina. 1992. The Groundworks of Danish Intonation. Copenhagen: Museum Tusculanum Press.

Jacobsen, Birgitte. (forthc.): The Question of 'Stress' in West Greenlandic. An Acoustic Investigation of Rhythmicalization, Intonation, and Syllable Weight.

Kotsinas, Ulla-Britt. 1987. Immigrant Children's Swedish - A New Variety? In: 2 Sprog, 2 Kulturer. Papers from the Fifth Nordic Conference on Bilingualism. Clevedon: Multilingual Matters. pp 129-140.

Kotsinas, Ulla-Britt & John Helgander (utg). 1994. Dialektkontakt, språkkontakt och språkförändring i Norden. Föredrag från ett forskarsymposium. Stockholm, MINS 40

Kotsinas, Ulla-Brit, Anna-Brita Stenström & Anna-Malin Karlsson (utg.). 1997. Ungdomsspråk i Norden. Föredrag från ett forskarsymposium. Stockholm, MINS 43.

Kristiansen, Tore. 1992. Har regionalsprog en fremtid i Danmark? In: danske folkemål, 34, pp. 203-247.

Lainio, Jarmo. 1989. Spoken Finnish in Urban Sweden. Uppsala Multiethnic Papers 15. Diss.

Langgård, Per. 1992. Grønlandsksproget - tosproget - grønlandsksproget: nogle tendenser i det dansk-grønlandske sprogmøde blandt Nuuk's skolebørn. In: Grønlandsk kultur- og samfundsforskning 92. Nuuk: Ilisimatusarfik/Atuakkiorfik, pp. 104-128.

Mase, Hideo. 1973. A study of the role of syllable and mora for the tonal manifestation i Greenland. In: Annual Report of the Institute of Phonetics, University of Copenhagen, Vol. 7. pp. 1-98.

Nagano-Madsen, Yasuko. 1993. Phrase-final Intonation in West Greenlandic Eskimo. In: Working Papers 40, Lund University, Dept. of Linguistics, pp. 145-155.

Nagano-Madsen, Yasuko & Anne-Christine Bredvad-Jensen. 1995. Analysis of intonational phrasing in West Greenlandic Eskimo reading text. In: Working Papers 44, Lund University, Dept. of Linguistics, pp. 129-144.

Petersen, Inge Lise. 1994. Dialektkontakt og regionalisme. In: Kotsinas & Helgander (utg.) pp. 242-251.

Rischel, Jørgen. 1974. Topics in West Greenlandic Phonology. København: Akademisk Forlag.

Trondhjem, Frederikke Blytmann. 1998. En sprogundersøgelse af en såkaldt halvsproget person. (This volume).

Wande, Erling. 1991. Sverigefinnars två språk. Språkbruk och attityder hos två generationer - en projektpresentation. In: Berge & Kotsinas (utg.), pp. 145-162.

Aasheim, Stine. 1997. "Kebab-norsk" - fremmedspråklig påvirkning på ungdomsssspråket i Oslo. In: Kotsinas et al. (utg.), 235-242.

# Meanings of the Word "Work" in a Danish Unemployment Office

*Helen Korsgaard Maxwell*

**Abstract**
In this paper I will discuss the word "work" with a point of departure in Michel Foucault's notion of discourse. He has described historiy as a sort of sequence of discontinuous historical discourses. He suggests that every individual subject is trapped in his or her own historical discourse with no access to other discourses. First I will introduce three encyclopedia entries, each from a different historical period. Secondly, I will discuss one of a number of conversations I filmed in an unemployment office in Denmark in 1992 and relate this to Foucault's historical understanding of discourse. Finally, I will offer an alternative analysis of the discourses I have found in my field research. Here I argue that our historical period is not characterized by a single master discourse, but by a variety of discourses which emerged in different historical periods.

Michel Foucault has described history as a kind of sequence of discontinuous historical discourses. He envisions every individual subject as trapped within his or her own historical discourse with no access to other discourses. In this paper I will examine the word "work" – "arbejde", in Danish – taking my point of departure in this perspective. First I will introduce three encyclopedia entries, each from a different historical period. Secondly, I will discuss one of a number of conversations I filmed in an unemployment office in Denmark in 1992, attempting to relate this to Foucault's historical understanding of discourse. Finally, I will offer an alternative analysis of the discourses I have found in my field research.

## 1. The word "work" in history

### *a. Religious discourse*

The word work – "arbejde" – has been defined publically many times in more recent Danish history. The first definition of the word I will examine is taken from the *Conversationslexicon* published in 1816. Here we may read that:

> *Work is any moral, permitted spiritual or physical effort to maintain or improve life, the goal of which is not immediate sensual enjoyment. Work is to live until Death releases the working man to a better life. The Earth cannot do without work, if it is to bear fruit. It is fair enough to repay him who sows wheat with wheat, and the lazy one with weeds. That is the principle of all work and justice. Man needs this in order to live and find joy in life..*[1]

In 1816 work is an effort that man must make in order to earn a living. The lazy, we are advised, will not reap any harvest. But work is not defined as a mere activity alone. Work is also a part of a religious view of the world – a view that relates work to the concept of justice. The worker works not only for himself, in order to earn his pay. He also works because he is a part of a larger moral system which demands that he work. Work is a fundamental part of life, closely tied to the soil. How well a man plows his field is connected to a larger sense of justice, in which the industrious are rewarded and the lazy are punished. I would tend to position this particular meaning of the word work within what we could generally refer to as a *religious discourse*.

### *b. wage earner's discourse*

In 1927 the concept has already changed considerably. I quote excerpts from the *Ordbog over det Danske sprog*:

---

[1] Arbeid er hver sædelig, tilladt Aands- eller Legemsanstrængelse til Livets Vedligeholdelse eller Forbedring, uden at denne Anstrængelse umiddelbart har sandselig Nydelse til Øjemeed. Arbeide er at leve indtil Døden afløser den Arbeidende til et bedre liv. Jorden kan ei undvære Arbeid, naar den skal bære frugter; den er retfærdig nok til at gjengive Den Hvede, som såede Hvede og den Dovne Ukrudt. Det er alt Arbeids og Retfærdigheds Princip. Mennesket behøver det for at leve og have Glæde af Livet. *Conversations Lexicon*. 1816. forelagt af A. Soldin og trykt af Græbe.

> *hard excercise of bodily or spiritual faculties, to attain a particular result or engage in an activity which demands a certain amount of effort. (the opposite of rest or play). Work done well produces wonderful fruit ... about economic activity which is carried out for another person. Wage work, the beginning state of conflict between capital and labour.*[2]

In 1927 work is still an effort, but it is no longer part of any religious understanding of the world. Nor is justice any longer connected to the concept of work. Work is an activity which people are paid to perform for themselves or, most often, for others. The day, this entry suggests, is divided into three parts: work, rest and play. These are the three main activities in which human beings engage. I would position this new sense of the word work in what we might call the *discourse of the wage earner*.

### *c. Psychological discourse*

In *Den Store Danske Encyclopædi* from 1994, work is introduced as an activity in which people engage in order to transform nature and thereby satisfy their needs. This definition, however, quickly becomes more complex. Aside from work, two other important human activities exist - play and learning. According to the encyclopedia these three activities are, in general, easy to separate, since work is directed towards the goal of transforming nature, while play involves activities that do not have a specific goal. The goal of learning is to develop and transform individuals themselves, not their surroundings. One difficulty that arises in trying to view these activities as separate, however, involves the fact that they, *'are interwoven. Work can be like play. Learning can be hard work, and playing can be part of one's work.*

---

[2] 1) *besvær; møje; genvordighed.* 1.1) *anstrengelse; besvær; møje; "slid". (ofte med overgang til bet.* 2). 2) *(anstrengende) brug af legemlige ell. aandelige evner, hvorved et vist resultat ell. virksomhed, som kræver en vis anstrengelse. (mods.* Hvile *ell.* Leg*).* 2.1) *godt Arbeide giver herlig Frugt [...] om erhvervsvirksomhed, som udføres i en andens tjeneste; Lønarbejde.* Det begyndende modsætningsforhold mellem kapital og arbejde. *Ordbog over det Danske Sprog*, 1919.

*Moreover, you can learn much by working or playing.*[3] Although work is an external activity, involving a transformation of nature itself, while learning is an internal activity which transforms those who learn, in 1994 the two processes appear interrelated, making the definition of work a more complex one than that which appears in 1927. I will position this most recent understanding of the word work within what might be termed as a *psychological discourse*. Each of these three dictionaries offers a description of at least one public understanding of the word work in more recent history. My description of these definitions implies that the meaning of the word work has undergone considerable change during this period – an observation which, I might add, is fully consonant with Foucault's understanding of history as a sequence of discourses. In the early 19th century, work is part of a religious belief system that also involves notions of morality and justice. At the beginning of the 20th century this has changed: work has been reduced to a mere activity in which people engage for others or for themselves. This understanding of the work is clearly connected to the rise of the wage earner as a consequence of industrialization – which began relatively late in Denmark. I have thus chosen to name the discourse to which this meaning of the word is tied *the discourse of the wage earner*. This meaning has been expanded again in our own day into a complex concept that interweaves work with nature and work with oneself. Work involves not only material production, but also – and I quote from the encyclopedia – "symbolic, spiritual and artistic production". Work is not only an external activity, but also an internal one, a notion which locates this meaning of the word work within a discourse of a more or less *psychological* nature.

---

[3] Virksomhederne væver sig ind i hinanden. Arbejde kan gå som en leg. Det kan være hårdt arbejde at lære noget, og det kan være en del af ens arbejde at lege. Desuden lærer man meget som biprodukt af arbejde eller leg. *Den Store Danske Encyclopædi*. Gyldendal, Nordisk Forlag, 1994: 529.

## 2. Description of two discursive systems underlying the word work in a Danish unemployment office

I have chosen to analyze a single conversation between an unemployed worker and an unemployment office counselor, Bent, who disagree over the value of education in applying for jobs. I will suggest that this disagreement stems from the fact that the worker and counselor base their respective views on two entirely different understandings of the word work – in fact, on two different discourses. The disagreement will thus appear as a clash between two discourses – discourses which, I will argue, turn out to be the "discourse of the wage earner" and the "psychological discourse" exemplified in the encyclopedia entries I have already discussed.

In this conversation the counselor predicts that tomorrow's labor market will make room for very few unskilled workers:

> Bent*: "... I am not saying that unskilled workers can't get a job. Very few unskilled workers will be able to get a job, but those who do will have a fairly good education.*[4]

Thus even if you are not specifically trained to perform the particular job you are applying for, any kind of education will be better than no education at all.

When the unemployed worker informs Bent that he has, in fact, found a job for which a worker trained to perform the job was turned down, Bent continues: '*Unfortunately, you mentioned that you got a job which a guy who was trained for it didn't get....*'[5] Bent finds it unfortunate that Arne can acquire this kind of experience – presumably because it contradicts Bent's understanding of the realities of today's labor market. Arne then takes issue with Bent's use of the word 'unfortunate' about his getting the job. 'Arne*:*

---

[4] Bent: Jeg siger ikke, at der ikke bliver nogen ufaglærte, der bliver ganske få ufaglærte og de ufaglærte, der bliver, de vil have ret høj uddannelse.

[5] Bent: ...desværre nævnte du, at du havde slået en der var mejerist i opløbet før...

*"Why are you saying unfortunate. I just had more experience than he had.*[6] Arne believes that the employer who hired him values experience over formal training. Therefore, what counts in the Danish labor market, according to Arne, is not education but experience. Arne believes that you will acquire all the education you need by actually performing the job. When Bent asks him about formal training, Arne expresses little faith in this:

> Bent: *"Have you taken any courses?*
>
> Arne*: No*
>
> Bent*: No courses at all, pp well yes?*
>
> Arne*: Are you talking about any courses in particular?*
>
> Bent*: No, no, I was just wondering if you took anything at the "specialarbejderskolen" or anything like that?*
>
> Arne*: No I don't buy that*
>
> Bent*: You don't buy that?*
>
> Arne*: No, all that'll do is make it take longer to get a real job.*[7]

Arne thinks courses are a waste of time. It is on the job where you acquire technical skills, which are precisely what Arne has learned is required in order to get a particular job. He is not interested in schooling.

As firm a believer in education as Bent is, he does not believe that technical training alone is enough to get a job. A worker also needs human insight. This human insight can be gained, for example, at a folkehøjskole:

---

[6] Arne: Hvorfor siger du desværre. Jeg havde da bare mere erfaring end ham?

[7] Bent: Hvad har du nogle kurser ellers?
Arne: Næ.
Bent: Slet ikke nogen kurser overhovedet pp godt ja.
Arne: Tænker du på nogen specielle kurser?
Bent: Nej, nej, Jeg tænkte bare på om du havde et eller andet fra specialarbejderskolen eller sådan noget?
Arne: Nej, den hopper jeg ikke på den der.
Bent: Den hopper du ikke på?
Arne: Nej, det er jo bare at forlænge det, at man render rundt der indtil man finder noget arbejde.

> Bent: *but then afterwards, when you have taken that, then you have to go to a Folkehøjskole.*
>
> Arne*: a folkehøjskole?*
>
> Bent*: a folkehøjskole, yes*
>
> Arne*: and do what? learn gymnastics?*
>
> Bent*: to be there, to be there*
>
> Arne*: to be there?*
>
> Bent: *yes*
>
> Arne*: just to be there. But there has to be a point in being there at a folkehøjskole. It can't just be to be there.*
>
> Bent*: no, that's because, no of course, it's just because a folkehøjskole gives you something.*
>
> Arne*: what?*
>
> Bent: *just p wider horizons , to put it really simply. You'll gain more insight, human insight, because you'll meet a lot of people there, they'll come from all kinds of different places and have a lot of different educational backgrounds and different experiences and skills.*[8]

Formal education is not merely training to perform a specific job. Other kinds of knowledge and qualities, such as "human insight" and "wider horizons" are also necessary. The counselor advocates a somewhat expanded notion of work when he suggests that more general wisdom such as human insight and wide horizons is required in addition to specific work-related skills. This additional wisdom or these inner qualities are consonant with the definition of work given in *Den Store Danske Encyclopædi* (cf. note 3). Arne on the other hand finds that good work requires practical skills that one acquires on the job, rather than in the classroom. If you spend time on formal training, you should minimize it as much as possible, and get on to the real thing: a job.

---

[8] Bent: ... men så bagefter, når du så har taget det, så skal du lige en gang på en øhh højskole
Arne: en højskole?
Bent: en højskole ja
Arne: og hvad og lære gymnastik?
Bent: og være der og være der.
Arne: og være der?
Bent: ja.
Arne: og være der bare for at være der der må da være et formål med at jeg skal på højskole, det kan da ikke bare være for at være der, altså.
Bent: nej, det er fordi, nej det er klart, det er fordi simpelthen, at en højskole den giver noget.
Arne: Hvad?
Bent: bredere horisont p simpelthen, hvis vi skal sige det enkelt, du får større indsigt, menneskelig indsigt, fordi du der vil møde en masse andre, der kommer med andre steder fra med andre uddannelser med and med en anden erfaring (475-482).

Arne's understanding of the word "work" fits into the *wage earner's discourse* I described earlier. That discourse emerged in the beginning of this century with industrialization and thus belongs, historically speaking, to the past. Arne and Bent – worker and unemployment counselor – thus stand, in a sense, for two different discourses. A Foucauldian reading of this problem might conclude that two orders of discourse underlie the meaning of the word work as it appears in this conversation, and I will now summarize the basic features of each of these two orders:

Arne (and ODS):
concrete experience on the labor market
technical skills
goal oriented knowledge
a wage for work done

Bent (and SDE):
psychological insight
personal development
social skills
formal education
holistic perspective

I have argued that the difference between these two discursive orders is the main reason for the clash that takes place between the client and the counselor in the unemployment office. It is interesting that the two orders can be traced back to two different historical periods. Arne the worker's discourse is highly akin to the *discourse of the wage earner* which emerged during industrialization and is described in the *Ordbog over det Danske Sprog*. Bent the counselor's discourse – which, it should be noted, represents official Danish labor market policy – involves both job-specific skills and more personal, inner qualities. This expanded sense of the word work is precisely what characterizes the entry on "work" in the new *Store Danske Encyclopædi*. I mentioned at the beginning of this talk that in 1816 the notion of work belonged to a larger religious belief system. Today, the word 'work' is connected, in part, to inner qualities of individual human subjects. Meaning in 1816 was a gift from God, while today it derives from the human psyche. Between these two understandings of the word "work" – both of which involve something more than mere applied effort – work appears to have been demythicized and, like in Arne's talk, involves little more than labor alone. In history, work appears to have changed character as it moved from the soil into the factories and from the factories into our psyches. But, interestingly, two separate "historical" understandings of the word "work" also seem to be operating at one time, in one place – in a single conversation.

## 3. An alternative analysis of discourse

In the first part of this paper, I outlined a brief analysis of the discourse of work from the historical pespective that Foucault elaborated. Yet the conversation I discussed above suggests that two men engaged in conversation in 1992 were informed by understandings of work which derive from two different historical periods. The unemployed worker Arne argues that work is the way it has always been (by which he means unadorned labour) and involves little more than applied efforts. The counselor Bent is informed by a discourse which appears highly contemporary, belonging to the post-industrial age. The counselor's discourse features a certain historical perspective: work was once performed by unskilled laborers, but today's work force requires a general education and various social and more personal skills. The counselor perceives work historically as a development away from how the worker Arne experiences it.

Bent's historical perspective can be described as giving him simultaneous access to several discourses. He can acknowledge Arne's discourse, but can also reject it as belonging to the past. Arne, on the other hand only seems to have access to his experiences and the discourse that describes them. I would therefore suggest that the clash between the worker and the counselor is not only a clash between two different discourses, but also a conflict between two individuals who do not have access to an equal range of discourses.

This leads me to suggest that Foucault's notion of a master discourse for each historical period – a notion supported by my examination of encyclopedia entries – runs into some trouble in light of my data material. For, the worker Arne's discourse seems to have emerged in the beginning of the century, but it is still alive today, right alongside a discourse which emerged in response to later labor-market and general historical developments. The fact that the counselor Bent is able to consider the worker's discourse – only to reject it – is further evidence that discourses originating in different historical periods operate simultaneously or are put to use in everyday interaction. Arne and Bent suggest, at least to me, that our present historical period is not characterized by a single master discourse, but by a variety of discourses which emerged in different historical periods.

*University of Southern Denmark*
*Main Campus: Odense University*
*e-mail: korsgaard@language.sdu.dk*

**Literature**

Conversations Lexicon, 1816. forelagt af A. Soldin og trykt af Græbe.
Foucault, Michel. The Archaeology of knowledge, 1972. Pantheon Books, New York.
Ordbog over det danske sprog, 1919. Gyldendalske Boghandel, Nordisk Forlag.
Den Store Danske Encyclopædi, 1994. Gyldendal Nordiske Forlag.

# Grices kooperative principper i kritisk belysning:
## Non-kooperation som diskursiv strategi i asymmetriske diskurser.

*Hans Jørgen Ladegaard*

**Abstract**
This paper discusses the notion of Gricean cooperation. In Grice's famous essay, and some of the pragmatic literature, it is assumed that human communication is rational and logical. It is argued that communicative interactions are cooperative efforts where speakers are oriented towards a common purpose or goal, i.e. to make the interaction succeed, even though this may sometimes appear not to be the case. However, examples from recordings of student-teacher interactions show that non-cooperation may be employed, not as a rare exception, but as the preferred discursive strategy. It is argued that we should distinguish between 'linguistic cooperation' and 'social goal-sharing', and finally, it is suggested that the notion of cooperation should be analysed in terms of what people want to obtain by their linguistic behaviour.

### Introduktion

Paul Grices lille artikel, 'Logic and conversation' (1975) er næsten blevet en klassiker inden for pragmatikken. Enhver pragmatik-tekstbog med respekt for sig selv har et afsnit om kooperation, hvor Grices artikel som regel danner udgangspunkt for diskussionen. I det følgende vil jeg kort skitsere hovedindholdet i Grices teori.

Udgangspunktet for Grices artikel er det kooperative princip og de deraf følgende maksimer, som siger, at menneskers kommunikative adfærd er rationel og logisk, og at dette kommer konkret til udtryk i den måde, hvorpå vi organiserer diskurser. Grice siger bl.a.:

> our talk exchanges do not normally consist of a succession of disconnected remarks, and would not be rational if they did. They are characteristically, to some degree at least, cooperative efforts; and each participant recognizes in them, to some extent, a common

> purpose or set of common purposes, or at least a mutually accepted direction. (1975: 45)

Denne hovedregel for den kommunikative praksis har Grice konkretiseret i et såkaldt kooperativt princip samt en række tilhørende maksimer (se appendiks).

Grice postulerer videre, at der som hovedregel *er* samarbejde til stede i en vikårlig diskurs - og at det er empirisk påviseligt - selv når det på det formelle plan ser ud som om, der *ikke* er samarbejde. Han siger bl.a.:

> ... it is just a well-recognized empirical fact that people DO behave in these ways; they have learned to do so in childhood and not lost the habit of doing so; and indeed it would involve a great deal of effort to make a radical departure from the habit. It is much easier, for example, to tell the truth than to invent lies. (1975: 48)

Derefter nævner Grice en række eksempler på, at der under specielle omstændigheder kan ske brud på ét eller flere af de kooperative principper, men han understreger, at et sådant brud, der *tilsyneladende* er et udtryk for en ikke-kooperativ handling, alligevel på et dybereliggende plan er et udtryk for samarbejde. Som et eksempel på denne dybereliggende samarbejdsvilje nævner han følgende interaktion (Grice 1975: 54). A siger: ‘Mrs. X is an old bag’, hvortil B svarer: ‘The weather has been quite delightful this summer, hasn't it?’. Indikationen i Bs tilsyneladende irrelevante svar er ifølge Grice, at As kommentar af én eller anden grund - måske fordi A ved sin bemærkning har begået en fadæse, eller fordi mrs. X uventet er dukket op - ikke skal kommenteres. Derfor fortolkes Bs respons som et udtryk for samarbejde på et *dybereliggende niveau*, selvom der tilsyneladende er tale om et brud på relevans-maksimet.

Hvis vi nu vender os til faglitteraturen inden for pragmatik, så synes tendensen - i hvert fald i de udvalgte værker jeg har kigget på - at være den, at der nok rejses en række kritiske bemærkninger til Grices teori, men at der sjældent sættes spørgsmålstegn ved det overordnede kooperative princip. Generelt er udgangspunktet det, at den kommunikative interaktion *er* rationel og logisk, at formålet (altid) er udveksling af information, og at vi per definition samarbejder om at få diskursen til at lykkes. Lad os kigge på et par eksempler fra litteraturen:

Stephen Levinson (1983: 103) argumenterer således for, at maksimerne kan have universel karakter. Han tager udgangspunkt i en række konkrete, ikke-sproglige eksempler på (non)kooperation og siger bl.a.:

> In each of these cases [examples of non-linguistic, non-cooperative behaviour] the behaviour falls short of some natural notion of full co-operation, because it violates one or another of the non-verbal analogues of the maxims of conversation. This suggests that the maxims do indeed derive from *general considerations of rationality applicable to all kinds of co-operative exchanges*, and if so they ought in addition to have *universal application*, at least to the extent that other, culture-specific, constraints on interaction allow. Broadly, this too seems to be so. (min emfase)

Georgia Green (1989: 98) antager samme rationelle, logiske adfærd hos interaktørerne i en diskurs og siger, at hvis ét af maksimerne brydes, så skyldes det, at diskursen anskues i et mere overordnet perspektiv, og at der således alligevel er samarbejde på det indirekte plan. Hun siger bl.a.:

> In most cases where a speech act appears not to conform with all of the maxims, the addressee is intended to assume that it IS in

> conformity with all of them, and interpret it in that light. In cases of genuinely disregarding a maxim, however, the addressee is intended to infer that the speaker is not conforming to a maxim, and figure out why.
> ... if *the speaker is assumed to be rational and cooperative, they* [seemingly non-cooperative remarks] *will be taken as cooperative with respect to some broader, more global enterprise involving speaker and addressee*, for example, maintaining appearances ... or just keeping a private conversation from being overheard. (min emfase)

Og Joseph Capella (1995: 172) argumenterer på tilsvarende vis og siger, at kommunikative interaktioner ville være at betegne som latterligt ineffektive - og i sidste ende gå i stå - hvis ikke det kooperative princip blev anvendt som normen. Han siger bl.a.:

> The force of Grice's insight is to show that *conversations would grind to a halt were the cooperative principle not employed as the norm.* If conversationalists did not follow the cooperative principle, then many interesting, subtle, and artful conversational exchanges would be rendered incoherent and would require so much elaboration as to give conversations the appearance of user-unfriendly computer manuals. The principle of cooperation can certainly be exploited ... but the point is that there exists an interpretive bias toward cooperation that is able to be exploited. Without such a principle, ordinary human interactions would be forced to be *ridiculously inefficient* ... (p. 172) (min emfase)

## Præsentation af data

Med baggrund i denne kortfattede indføring i begrebet 'Gricean cooperation' vil vi dernæst kigge på en række diskurser - mere præcist interviews/lærer-elev interaktioner - hvor man kan stille spørgsmålstegn ved det paradigme, der netop er blevet skitseret, altså at den kommunikative interaktion fremstilles som en rationel, logisk og kooperativ handling, hvor samtalepartnerne arbejder sammen om at nå ét eller andet fælles mål.

Følgende eksempler er uddrag af interviews i to 9. klasser i hhv. en rural og en urban kontekst. I den rurale kontekst (eksempel 1 og 2) skal eleverne forberede deres erhvervspraktik, og læreren udspørger dem om deres forventninger til praktikken samt deres ønsker vedrørende uddannelse og arbejde generelt. I den urbane kontekst (eksempel 3) arbejder klassen med temaet skole og uddannelse i samfundsfag og skal i den forbindelse besvare en række spørgsmål om emnet. Optagelserne er foretaget som led i et eksperiment, hvor formålet var at elicitere sprogholdninger og sproglig adfærd hos samme informantgruppe. Det betyder altså, at interviewene ikke eksplicit havde til formål at belyse det, der er genstand for diskussion i denne artikel, nemlig kooperation. (. , .. , ... etc. angiver pauser af omtrent 1, 2, 3 etc. sekunders varighed).

### Eksempel 1

**Informant nr. 38 (Klaus) og AS (lærer, K)**

(1) AS: Klaus .. nu har du chancen for at bruge munden
K: Uha det ka' a" (latter)
AS: Hva' siger du du har skrev'n
K: A hår æ" anden skrøwn jen ting det ve' a faktisk æ" det war æ tømrer op' i Tøndering
(5) AS: Ja ... det vil altså sige at du har egentlig ... har du prøv' på selv at find' noget
K: Nej det we' a egentlig æ"

AS: Hvem har så funden det til dig så vidt jæ ka' husk' så stod der da både det det var da arrangeret

K: Jo jo men det var fur at mur hun hår æ" (uforståeligt) det vil' a da godt

AS: Ja

K: så: ring' hun derop

AS: Ja

K: og det ku' a så godt

AS: Ja ... det der med tømrer Klaus er det det du går og drømmer om .. //altså ku' du godt tænk' dig// at bliv' håndværker

K: //det ved a såmænd æ"// det ku' vær' sjov nok og prøv'

AS: Ja ... det vil sige alt det her du bruger jo utrolig meget tid på landbrug selvfølgelig er det egentlig ik' det du vil

K: Jo det trower a da . det ka' a altij find' uk' å

AS: Hva'

K: Det ka' a altij find' uk' å

AS: Det ka' du altid find' ud af det er rigtig nok . men det vil sige jæ synes da også det er fornuftigt nu vil du så prøv' det her .. er der var der andre håndværk du ku' tænk dig at prøv'

K: Mmm det we a såmænd æ" .. nej det war der wel æ"

AS: Nej det sku' være noget med træ så hvis det var sådan ... du sku' prøv' noget

K: [pause, 5 sek.]

AS: godt .. er det der et sted du kender

K: Ja a we' da hwor den leger hen' men ellers så .. we a da"

AS: Du kender æ" æ folk

K: Jo a kender ham sjal'

AS: Er det en stor .. virksomhed eller er det bare **en** mand du ska' med

K: A we æ" rigtig hwor manne der er der ...

AS: Nå [pause, 6 sek.] Er der langt dertil

K: Næ de:t ... der er wo:l en fem kilometer derop eller så'en nøj

AS: Ja ..det vil sig' do ska' på æ cykel

K: Ja det we a" rigtig hwor'n vi finder uk' o'

AS: Ja hvordan war æ no mæ det hæe hvis do sku' cykel .. hvis vi nu kigger på transport til det her praktik hvordan var æ så man gjorde det .. hvis nu du cyk'ld hwa så

K: Ja så cyk'ld a da (latter)

AS: Ja . kun man få noget ud af det

K: Det ve' a" . det ve' a" . æ" and'n motion

AS: Jæ ka' sørme ik' husk' det var ik' ret mange ører du ku' få men du ku' få nogen få ører pr. kilometer hvis du var over så'en den der cykelgrænse der er .. så det må du så prøv' og tænk på men til Tøndering der er der nok ik' . er der offentlige transportmidler den vej
K: Det we a"
AS: Går der bus
K: Det trower a"
AS: Det trower du ik' . du har æ" lav't de store undersøgelser endnu så
K: Nej .. det er wist nemmest og ta' æ cykel
AS: Hva' siger du
K: Det er wist nemmest og ta' æ cykel
[...]

## Eksempel 2

**Informant nr 40 (Thomas) & AS (lærer, K)**

AS: Thomas .. hva' siger du Thomas [pause, 5 sek.]
T: (svagt) Det wed a et
AS: Det ved du ik' ... du plejer da ik' være så ked af at sige noget da i hvert fald
T: (svagt) Hva' .... nej a hår valgt hva' hår a valgt .. æ militær og så øh .. elektriker og .. det sidst' //
AS: //og hva' sagde du til sidst
T: (svagt) A så'e .. det ska' a lige tænk' over føst . nåe ja radio og: . radioTV . og så øh .. kom' u' og sejl' . eller det er nøj der er fast .... wor praktik // (uforståeligt)
AS: //prøv lige og snak' lidt højere// hva' sagde du
T: (højt) Det regner wi i hvert fald maj ... (svagt) at a ska' u' og sejl' (uforståeligt)
AS: Altså det vil sige at du altså du lægger an på nu her du ska' ud i praktik og så vil du prøve og søge og finde ud af om om det ka' lade sig gøre at komme ud og sejle
T: Jo .. eller (uforståeligt)
AS: Har du snakke' med Kjeld om det egentlig
T: Nej
[...]

### Eksempel 3

**Informant nr. 10 (Thomas), nr. 11 (Christel) og CH (lærer, M)**

CH: Thomas jeg ku' godt tænke mig at spørge dig .. på flere efterskoler har man kørt forsøg med en erhversforberedende elvte klasse .. kunne du tænke dig at komme i elvte klasse
T: Nej ..
CH: Hvof//
T: //fordi jeg gider ikke gå i skole mere .. efter tiende [pause, 5 sek.]
CH: Gider du ik'
T: Næ
CH: Hvorfor gider du ik' det
T: Næ
C: Nu har du gået i skole i ti år så://
T: // elleve [pause, 5 sek.]
CH: Hvorfor gider du ik' det
T: Fordi jeg synes det er lidt rigeligt med tiende . når jeg er færdig med tiende .. så' jeg været lang tid i skole ... elleve år ..
[...]

CH: Thomas du sidder og snakker hele tiden [pause, 5 sek.]
T: Ja .. hva' skal jeg sige
CH: Ja sig noget
T: Hej (latter)
CH: Er der ik' noget øh du har tænkt over af det her . hva' glæder du dig til at forlade folkeskolen
T: **Ja**
CH: Nå .. det vil jeg da godt ha' en forklaring på
T: Nå ... ja så øh: fri for lærerne (latter) ... jeg gider it' sidde her hele mit liv
CH: Så skal du ud og lave noget jo
T: Det er da og' sjovere
[...]

Lad mig sige allerførst at disse eksempler på, hvad man kunne kalde 'non-kooperativ adfærd', ikke var vanskelige at finde i mine lærer-elev interviews. Tværtimod, de forekommer så hyppigt i mit materiale, at de hos nogle informanter fremstår som normen mere end undtagelsen. Man kan naturligvis sige mangt og meget om disse eksempler, men jeg har valgt at kommentere en række observationer, der relaterer sig til begrebet kooperation.

(1) I eksempel 1 er det først og fremmest påfaldende at Klaus ikke udviser nogen form for lingvistisk konvergens. Ifølge teorierne - 'Speech Accommodation Theory' (Giles 1973) og 'Audience design' (Bell 1984) - burde Klaus konvergere lingvistisk i retning af standardsproget for, som socialpsykologerne ville udtrykke det, at opnå samtalepartnerens sociale anerkendelse, altså som et udtryk for samarbejde. Dette sker imidlertid ikke, hvorimod AS konvergerer (se linie 30, 36, 38-39 og 49), selvom man ifølge teorien kunne formode, at det ville være Klaus, der p.g.a. sin underlegne sociale position, ville være mere tilbøjelig til at udvise lingvistisk konvergens.

(2) Hvis vi dernæst prøver at kigge nærmere på Grices maksimer i relation til Klaus og AS's diskurs, så ser vi igen den samme tendens fra Klaus, altså mangel på kooperation. Et oplagt eksempel er Grices første maksim (Quantity: 'Make your contribution as informative as is required', p. 45). Her er der flere tilfælde, hvor informanten bestemt ikke er så informativ, som det kunne forventes (linie 6, linie 16-17, 25, 27, 33 og 40). Vi kan også nævne Grices 2. maksim (Quality: 'Try to make your contribution one that is true', p. 46), som bliver overtrådt i linie 5-6, hvor lærerens spørgsmål '.. har du prøve' på selv at find' noget' bliver besvaret med et 'det we' a egentlig æ'". Logisk set må Klaus selvfølgelig vide, hvorvidt han har prøvet at finde en praktikplads eller ej, og svaret er derfor usandt. Det samme gælder for linie 46-50 angående transportmuligheder til praktikken. Der er tale om et meget lille lokalsamfund, og Klaus ved helt klart, om der går bus til pågældende praktikplads. Vi kan også nævne 3. maksim: (Relevance: 'Be relevant'). Dette princip bliver brudt i linie 7-9, hvor Klaus reelt set ikke besvarer det konkrete spørgsmål, han får stillet, nemlig 'Hvem har fundet din praktikplads'. Et tilsvarende eksempel er linie 38-42, hvor lærerens spørgsmål 'Hvis du cyklede, hvad så?' refererer til emnet cykelpenge og transportudgifter, som allerede er blevet indgående behandlet i klassen, men

hvor Klaus vælger at gøre sit svar til en vittighed ('Ja så cyk'ld a da'), og straks høster kammeraternes anerkendelse i form af latter.

(3) Vi kan gå videre til eksempel 2, hvor vi i indledningen af interviewet ser samme tendens til manglende kooperation. Igen ser vi en informant, der er meget lidt informativ (line 2: 'det wed a et'), og så snakker han tilmed utroligt stille, hvilket betyder, at AS ikke kan høre ham. Også dette kan ses som et udtryk for non-kooperation. Heller ikke Thomas konvergerer lingvistisk i retning af AS, men taler udpræget dialekt gennem hele interviewet. I sekvensen linie 10-11 ser vi et eksempel på, at nogle af maksimerne brydes, nemlig 3. 'Relation' og 4. 'Manner'. AS har ikke hørt hvad Thomas' praktikønsker går ud på og vil gerne have ham til at gentage, men Thomas vælger at gentage, højt og tydeligt: 'Det regner wi i hvert fald maj' (linie 11), hvilket selvsagt ikke er den oplysning AS ønsker gentaget.

(4) I eksempel 3 ser vi tilsvarende eksempler på en usamarbejdsvillig informant (se linie 25-33). Thomas' 'Næ' i linie 33 er et klart brud med både 'Quality', 'Relation' og 'Manner'. Tilsvarende i linie 40-43, hvor CHs anmodning om at Thomas skal sige noget - underforstået fornuftigt om det emne der bliver diskuteret - resulterer i et 'Hej', som ikke følger maksimerne mht. hverken kvantitet, kvalitet eller relevans, men som derimod resulterer i en tilkendegivelse af kammeraternes anerkendelse (latter, linie 43).

## Diskussion og konklusion

Et fællestræk for disse og mange tilsvarende eksempler fra mit materiale er, at de giver anledning til at sætte spørgsmålstegn ved det overordnede kooperative princip. Der er generelt ikke tale om samarbejde, hverken formelt eller på et dybereliggende plan, og der er heller ikke tale om, at disse interaktioner på ét eller andet plan er logiske og rationelle. Jeg mener vi kan konkludere, at på i hvert fald ét plan giver disse interaktioner eksempler på non-kooperativ adfærd. Et af hovedproblemerne i denne diskussion synes nemlig at være, at vi mangler en præcis definition af begrebet 'kooperation'. Jenny Thomas - som udtrykt hos Sirangi & Slembrouck 1992 - skelner i sin analyse mellem to forskellige former for kooperativ adfærd, nemlig:

(a) **'linguistic cooperation'** (et sprogligt/interaktionelt samarbejde hvor spørgsmål og svar passer sammen på det formelt lingvistiske plan), og

(b) **'Social goal sharing'** (interaktørerne samarbejder i henhold til givne sociale normer - fx for adfærd i klasseværelset - om det at løse en bestemt opgave).

I de ovenfor nævnte eksempler sker der kun samarbejde mht. (a) (der er ingen eksempler på 'Goddag mand økseskaft', idet spørgsmål og svar formelt passer sammen), men *ikke* mht. (b), idet eksemplerne netop viser, at interaktørerne *ikke* har et fælles mål for deres diskurs. Resultatet bliver derfor 'ridiculously inefficient human interactions' (jvf. Capella 1995), som fx ofte går i stå (jvf. de mange pauser).

Det betyder altså, at vi ikke uden videre kan tale om kooperativ adfærd uden at eksplicitere, hvilken form for samarbejde vi taler om. Jeg mener også, at vi med baggrund i de eksempler, vi her har kigget på, kan stille spørgsmålstegn ved universaliteten i det kooperative princip, og ligeledes om diskurser, der er non-kooperative, kan betegnes som 'marked'[1]. Problemet hos Grice - og måske ikke mindst hos nogle af de pragmatikere, der siden har diskuteret hans teori, så som Capella, Green og Levinson - er måske en alt for énsidig opfattelse af, *hvad* en diskurs er, eller hvad *formålet* med en diskurs kan være.

Peter Harder (1996) argumenterer således for en revision af Grices informationsorienterede forståelse af begrebet 'meaning' (altså at formålet med diskursen per definition er udveksling af information) og siger bl.a. (p. 133): 'Because of Grice's purely informational understanding of meaning, his account needs to be revised in the light of the *functional approach* to linguistic meaning' (min emfase). Et tilsvarende synspunkt kommer til udtryk hos Jacob Mey (1993: 75), som foreslår at

> ... the whole problem of 'flouting' a maxim should be rethought in terms of **what people want to obtain by their linguistic behaviour**, that is to say, in the light of the pragmatic aspect of cooperation.

Jeg vil argumentere for, at for Klaus, de to Thomas'er (samt alle de andre) er det et formål i sig selv at udvise en ikke-kooperativ adfærd. Jeg mener, vi kan se de tre informanters adfærd som et udtryk for, at de ikke *vil* interviewes, men at de samtidig befinder sig i en situation, hvor de ikke *kan* sige nej. De kan ikke direkte sige 'Det rager ikke dig!' eller 'Jeg gider ikke svare på dine spørgsmål', og derfor må de - på en langt mere indirekte måde - markere deres uvilje mod interviewet. Vi kan altså, mener jeg, betegne non-kooperation som en diskursiv strategi.

Jeg mener derfor, at vi må tilsidesætte tanken om universalitet i forhold til Grices teori og i stedet forsøge at gøre den *virkeligt pragmatisk* ved at argumentere for, at kun gennem en analyse af konteksten og situationen kan vi få et indblik i, hvorvidt der i en given diskurs er tale om samarbejde eller ej.

*Institut for Sprog og Kommunikation*
*Odense Universitet*
*Campusvej 55*
*DK-5230 Odense M*
*e-mail hjl@language.ou.dk*

**Noter**

(1) Man kan naturligvis vælge at se disse eksempler på non-kooperativ adfærd som 'markerede former', altså et udtryk for at maksimerne eksisterer og i disse specifikke tilfælde ikke bliver overholdt (jvf. Capella 1995: 172: 'the point is that there exists an interpretive bias toward cooperation that is able to be exploited'). I så fald må vi imidlertid også argumentere for, at de såkaldte 'markerede former' i denne kontekst - qua deres frekvens og deres status som den foretrukne diskursive strategi - ophører med at være markerede og i stedet synes at blive normen.

## Referencer

Bell, A. 1984. Language style as audience design. *Language in Society* 13, 145-204.

Capella, J.N. 1995. An evolutionary psychology of Gricean cooperation. *Journal of Language and Social Psychology* 14, 167-81.

Giles, H. 1973. Accent mobility: A model and some data. *Anthropological Linguistics* 15, 87- 105.

Green, G.M. 1989. *Pragmatics and the Natural Language Understanding.* Hillsdale, NJ: Erlbaum

Grice, P. 1975. Logic and conversation. In P. Cole & J.L. Morgan (eds.), *Syntax and Semantics* (vol. 3), 41-58.

Harder, P. 1996. *Functional Semantics.* Berlin: Mouton.

Levinson, S. 1983. *Pragmatics.* Cambridge: Cambridge University Press.

Mey, J.L. 1993. *Pragmatics. An Introduction.* Oxford: Blackwell.

Sarangi, S.K. & Slembrouck, S. 1992. Non-cooperation in communication: A reassessment of Gricean cooperation. *Journal of Pragmatics* 17, 117-54.

Thomas, J. 1995. *Meaning in Interaction: An Introduction to Pragmatics.* London: Longman.

## Appendiks

### Grice's Cooperative Principle:

Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged (Grice 1975: 45)

### The Four Maxims:

1. QUANTITY
(a) Make your contribution as informative as is required (for the current purpose of the exchange).
(b) Do not make your contribution more informative than is required.

2. QUALITY
(a) Do not say what you believe to be false.
(b) Do not say that for which you lack adequate evidence.

3. RELATION
Be relevant!

4. MANNER
(a) Avoid obscurity of expression.
(b) Avoid ambiguity.
(c) Be brief (avoid unnecessary prolixity).
(d) Be orderly.

(Grice 1975: 45-46)

# Codeswitching Between Persian and Swedish

## A syntactic approach

*Leila Naseh*

**Abstract**
In the past decades the migration flow to the industrialized European countries has created a mosaic of language use behavior. In Sweden for instance Finnish, Spanish, Turkish, Polish, Arabic, Persian, Russian and a number of other minority languages are spoken along with Swedish. Dutch and Arabic are spoken in Holland. Turkish and Italian are spoken alongside German in Germany and so on. The significance of this cultural and linguistic diversity has raised considerable scientific interest in studying the language contact situations, bilingualism and language use behavior of immigrant populations. As these studies suggest immigrant bilinguals within the potential linguistic repertoire accessible to their community and according to their bilingual competence may demonstrate different types of language use behavior. *Codeswitching* , the use of more than one language in the same utterance is one of these behaviors which is common among certain bilingual immigrant communities. Codeswitching as an important aspect of bilingualism has attracted a considerable amount of research from different linguistic perspectives including structural and sociolinguistic. This paper is also concerned with the codeswitching phenomena. The goal has been to examine which structural constraints govern codeswitching practices when two typologically divergent languages are involved. In this regard Iranian immigrant bilinguals and their codeswitching performance between Persian and Swedish provide an interesting case in the point. Persian is a language that has received little attention in the Nordic countries. I hope this investigation will be a useful contribution to cross-linguistic studies of language contact phenomena.

## 1. Introduction

The purpose of the paper is to examine the codeswitching phenomena from a syntactic and structural view point. The analysis of the study is mainly based on tape recorded naturally occurring intragroup

conversations of 20 young Iranian immigrant youths studying at the high-schools in Stockholm.
Investigation of language choice behaviors among this group suggests that in some domains and with some interlocutors( at home with parents and older members of the family), Persian operates as the dominant language of conversations while some single or two word phrases are switched from Swedish into Persian utterances.

A second pattern used by the young bilinguals involves alternation of the two languages or frequent switching from Persian to Swedish. This pattern is widely used by the teenager group in intra-peer group interactions where all the participants are similar bilinguals. The patterns of codeswitching common in the speech of the majority and among the young bilingual speakers are exhibited in the following examples:

1) vaqti ***irriterad*** hast-am suedi harf mi-zan-am
when irritated am-1sg Swedish speak IMPF-do-1sg present
‘ I speak Swedish when I am irritated ‘

2) Mixây be-gam chetori, genom att inte visa känslor för henne
want subjunctive-say how through that not show feelings for her
‘ Do you want me to say how, through not showing feelings for her ‘

3) ye ***medelklass familj*** tu-ye Iran ***i genomsnitt två bilar*** dâra-an
a middle-class family in- e ezaf Iran on average two cars have-3pl present
‘ A middle-class family in Iran has on average two cars ‘

Example (1) is the most common pattern of codeswitching across the bilingual members of the community. As we see in (1) a Swedish single word is inserted in Persian sentence. The second pattern exhibited in (2), is used by the teenager group in intragoup interactions. In (2), the Swedish infinite clause is embedded in Persian matrix by means of the Swedish PP. In (3) Swedish phrases are inserted in the Persian matrix sentence. As we note Swedish phrases retain their Swedish configurations, however they are inserted in syntactic positions compatible with Persian principles.

As the above examples exhibit direction of switches are predominantly from Persian to Swedish, that is Swedish elements are embedded in Persian matrices. Persian single words or multi-word fragments rarely occur in Swedish utterances. We can therefore conclude that Persian /Swedish codeswitching shows a unidirectional character.

## 2. Collecting the data

The young Iranian bilinguals participating in this study consisted of 20 young girls and boys at the age group 17-23. These young group of speakers were studying at the high-schools in Stockholm. The length of their residence in Sweden at the time of recording(1994-1996) varied between 8-10 years.

The conversations were tape recorded in the course of seven informal conversations among the youths.Each recording session lasted between 60-90 minutes. All speakers were fluent in both languages and were school friends.

## 3. The phenomena of codeswitching

The term codeswitching refers to the use of two languages in a single utterance. In such utterances some constituents may belong to one language and some to the other language.

The earliest definition of codeswitching dates back to Weinreich (1953), who gives a situational character to it and defines the bilinguals as an ideal individual who ' switches from one language to the other according to appropriate changes in speech situation '.

In more recent studies of codeswitching, it has been found that switching occurs even within a sentence. The term intrasentential codeswitching is used for switches within the sentences. The term intersentential codeswitching is used for switches between sentences. Intrasentential codeswitching is very common in language contact situation where bilingual speakers share a similar linguistic background. poplack (1980) suggests that codeswitching behavior may be used to measure bilingual ability.

## 4. Constraints on codeswitching

In the last two decades a large number of studies on codeswitching have been concerned with the intrasentential codeswitching and analysing how the grammatical and syntactic rules of the two language systems interact in codeswitched sentences. The question to be addressed by such analysis has usually been what structural principles govern the configuration of codeswitched sentences and what principles block switching to occur at certain points.

As a result of these studies it has been proposed that codeswitching is rule-governed and is constrained by either language specific, universal or theoretical constraints (Pfaff, 1979; Gumperz, 1979, 1982; Bentahila & Davies 1983; Joshi 1985; Nishimura, 1985, 1997; Di Sciullo, Muysken and Sing, 1986; Treffers-Daller 1990; Myers-Scotton, 1993; Blazi, Rubin and Turibio, 1994).
Formulation of different models of constraints on codeswitching in general fall into four major groups:

1) Language specific constraints
2) More general constraints
3) Theory based constraints
4) Matrix language approaches to the constraints

Testing the different models of constraints on Persian/Swedish codeswitching data has not showed unified results. Some cases of codeswitching satisfy the predictions of some models and other cases fall under the predictions of another.

Space does not allow examination of all the models and every boundary of switching. We therefore examine only the predictibility of a more general and two theory based constraints. Predictions of Myers-Scotton's Matrix Language Frame model of codeswitching will also be examined.

## 5. An overview of some models

The equivalence constraint (Poplack 1980, Poplack &Sankoff, 1981) which is claimed to be universal, is one of the most influential constraints on codeswitching which has received most attention in the literature and has attracted most criticism. Poplack, in proposing the equivalence constraint, asserts that some kind of equivalence between surface structures of the two languages at the switch point is required before a switch is possible.

Poplack (1980, 1981) further to the equivalence constraint has also proposed, the Free morpheme constraint. This constraint rules out switches between a free root morpheme and inflectional or derivational morphology. Later on the notion of equivalence at the surface structure which was given an informal formulation in earlier works (Poplack 1980,1981) was examined at more abstract levels.

Bentahila and Davies (1983) were one of the first who examined that more abstract level equivalences between languages involved in codeswitching account for switchability. These scholars proposed the subcategorization principle according to which " all items must be used in such a way as to satisfy the [language-particular] subcategorization restrictions imposed on them ".

Some other researchers inspired by the generative grammar of Chomsky (1981), then current in theoretical syntax, proposed syntactic conditions on codeswitching within the generative grammar. These researchers including Woolford (1983), Disciullo, Muysken and Sing (1986) explained codeswitching phenomena in relation to constituent structures formulated under the syntactic theory of government and binding (GB). Woolford (1983) proposed the phrase structure congruent model of constraints and Diciullo *et al.* proposed the government constraint.

Under Woolford's (1983) phrase structure congruence model of constraints, the two monolingual grammars cooperate in code-switching but none of the rules of the two languages are changed. Woolford suggest that, where the two languages share a common PS-rules, lexical items can be freely drawn from either language to fill terminal nodes created by PS rules common to both languages (1983:535). Where the PS-rules or the word orders are different switching is prohibited. Thus a sentence can be generated by applying any rule from the two languages, as long as they are compatible.

Although Woolford (1983) is one of the several researchers who attempts to explain the constraints within the generative grammar framework and in terms of GB, however her formulation of the constraint is an alternate to equivalence constraint.

Disciullo, Muysken and Sing (1986) developed the government constraint. According to DiSciullo *et al.*'s model switching is not possible when there is a government relation, unless the governed element includes a " language-carrier, Lq-carrier ", the language of which matches the language of the governing element. In another word the governor and the highest lexical element in the governed maximal projection have to have the same language index. According to this constraint no codeswitching occurs between a lexical head ( N, A, V, P) from one language and any phrase from another language when there is a government relationship holding between them.

Belazi, Rubin and Toribio (1994) on the other hand proposed the Functional Head Constraint. Belazi *et al.* follow Abney (1987) who posit that there exists a special relation between a functional head and its complement. This relation in Abney's term is called f-selection. Belazi *et al*. claim that the strong relation that exists between a functional head and its complement impose constraint on switching between such heads and their complements.

Belazi *et al*. (1992,1994), distinguish between the lexical heads (V,P) and the Functional Heads (quantifier, negation element, modal and INFL) and assert that switching between a functional head and its complement is blocked. In contrast to this restriction, they claim that switching between a lexical head (V,P) and its complement occurs quite freely. At this point Belazi *et al*.'s analysis differ totally from DiSciullo *et al*. (1986). For instance switching between V and its complement is a free switch point in functional head constraint, whereas according to DiSciullo *et al*.'s government constraint, it is blocked by the government relation that holds between them
Belazi *et al*. (1994) also propose a second principle on codeswitching: The Word Grammar Integrity Corollary. According to this principle a word obeys the grammar of its language. In proposing this principle Belazi *et al*. follow Chomsky (1993), that lexical entries are associated with morphological and syntactic features. They argue that since this principle holds up for all lexical items in all languages, thus it must be valid for codeswitching as well. The WGIC for instance requires that both the adjective and the noun should obey the grammar of their respective languages. This grammatical restriction similar to that proposed by Poplack in equivalence constraint also rules out switching between adjectives and nouns in languages with pre-and postnominal adjectives.

The matrix language approaches to the constraints ( Joshi 1985; Myers-Scotton 1993a) account for differentiating the role of the two languages in codeswitching.

Myers-Scotton(1993a) in Matrix Language Frame model of codeswitching, posits two principles:

a) The morpheme order principle and
b) The system morpheme principle.

The system morpheme principle predicts that:

> *The ML supplies all productive system morphemes (i.e., productive inflections and function words) in Ml+El constituents.*

## 6. Persian/Swedish counterexamples to the equivalence constraint

Interesting environments for switching between Persian and Swedish in violation of the Poplack's equivalence constraint were found at the boundaries where the two languages differed in their surface structure configurations. As was said before according to the equivalence constraint switching between the linguistic elements of the two languages would be impossible if the surface orders of these elements at the point where switching is to take place mismatch.

Two main body of switches occurring in the data corpus are indeed those prohibited by the equivalence constraint. These were switching between verb and objects on the one hand and between noun and adjectives on the other. Some counterexamples are exhibited in the following:

### *6.1 Switching between object and verb*

Clearly there is no surface structure equivalence between syntactic order of verbs and objects in Persian and Swedish. Persian is a head-last language and verbs take their complements to the left. Whereas Swedish is a head-first language and complements to the verb appear to the right of the verb. Yet switching between Persian verbs and Swedish objects occured very often. The data exhibited the examples as in (4&5):

4) ***Högskoleprov*** ***-et-*** o dâd-am
university exam-the-râ gave-1sg past
' I passed the university exam '

As we note in (4) a Swedish object NP is followed by a Persian verb. We also see that in (4) the Swedish object complement is double marked with both the Swedish definite marker the suffix *-et* and with the Persian definite/object marker, the particle *râ* (0/ro).
In (5) as below a Swedish infinitive verb and its internal preposition *på* ' to' is followed by a Persian oblique object complement:

5) vi måste, vi måste tänka på ***hovviat***
we must, we must think to identity
' We must, we must think to the identity '

### *6.2 Counterexamples to the Free morpheme constraint*

Switching at the morpheme boundaries internal to words are also another possibility that occurs very often in the corpus while the free-morpheme constraint (Poplack, 1980) blocks it. The free-morpheme constraint predicts that switching between a root morpheme and inflectional or derivational morphology is prohibited. However in my corpus in many cases Swedish predicative adjectives end up with the Persian copula *ast* 'is', or Swedish nouns or infinitive verbs end with the Persian clitic object pronoun *e_ / et* ' its/her/him' as in (6):

6) *anmäla-_* *kard-im*
report-clitic pro did-1pl past
' We reported him '

I also found that very often Swedish nouns were suffixed with Persian plural inflectiont the suffix -hâ, as well as with Persian -e ezâfe(-e ezâfe connects the head nouns and their referents) as in (7):

7) ***Livsmedel***-hâ-ye Täby
grocery-pl-Ez Täby
' Groceries in Täby '

## 7. Government & Functional head constraints:

The Government constraint model predicts that since subject is not governed by the verb, hence switching between subject NP and VP is free. The Swedish/Persian codeswitching data shows the following:

### *7.1 Switching between subject and verb*

Switching between subject NPs and VPs as government constraint predicts occurred quite often in my data. But switching between pronominal subject and verb was absent. The absence of switching between pronominal subject and the verb might be explained by the pro-drop character of Persian and that the properties of the dropped subject appear on INFL. Thus the empty category in Persian is governed by INFL.

The nonoccurence of switching between pronominal subject and verb is compatible with the predictions of Functional Constraint (1992). Belazi *et al*. claim that under the FHC, switching between functional heads such as INFL, NEG and their verb complement is blocked. These predictions correctly account for nonoccurrance of switching between Persian INFL and Swedish verbs, as well as for nonoccurance of switching between a NEG element and verb. Swedish verbs have always occurred in the bare infinitive forms and in combination with the Persian auxiliary verb *kardan*.
*Kardan* ' to do ' is inflected for tense, aspect, person and number in agreement with the subject.

### *7.2 Switching between verb and object*

Di Sciullo *et al*. (1986) predict that under the head-complement relations of X-bar theory, switching between a governing and a governed element is prohibited. They assert that a lexical governor and the highest lexical element of the maximal projection it governs must be in the same language. Since verbal complements and object NPs are governed by the verb according to the government model they must be in the same language as the verb. Hence switching between the verb and object NP on the one hand and between the complementizers and conjunctions on the other are not permissible.

As I showed above in (4) and (5) switching between a Persian finite verb and its Swedish complement object occurred quite often. Switching between a Swedish verb and its Persian object also occurred though far less than switching between a Persian verb and its Swedish object complement. As such predictions of government constraint at this switch boundary do not hold up for the data and we can conclude that switching between Vand NP is permissible in Persian/ Swedish codeswitching. This also implies that Belazi *et al.*'s prediction that switching between a lexical head(V, P) and its complement occurs freely, is compatible with Persian/Swedish data. In the following example a Swedish noun occurs as the complement of a Persian preposition. The following is also a counterexample to the government constraint:

8) bâ *kniv* bebor
with knife cut
' Cut with the knife '

### *7.3 Switching between COMP and IP*

Restriction of switching between a complementizer and IPs which is predicted by both Di Sciullo *et al.* (1986) and Belazi *et al.*(1994) is refuted by the data. The following is a counterexample to this restriction:

9) ne-mi- tun-an ke *undvika varandra*
not- IMPF can-3pl COMP to avoid each other
' They cannot avoid each other '

Both constraints are clearly ruled out in (9), since the language of the COMP is differnt from that of the verb.

### *7.4 Switching between noun and adjective: a counterexample to four constraints*

The equivalence constraint (Poplack, 1980), and the phrase structure congruence model (Woolford, 1986), prohibit switching between a noun and the modyfying adjective if the order of these constituents is

383

different in the two languages. The government constraint also predicts that the nouns governing the adjectives must come from the same language.

The Word Grammar Integrity Corollary presented by Blazi *et al*. (1994) accounts for placement of modifying adjectives in codeswitching if the grammar from which they are drawn match together. The WGIC therefore similar to equivalence constraint predicts that switching between adjectives and nouns in languages with pre and post- nominal adjectives is impossible. Since Persian adjectives are postnominal and Swedish adjectives are prenominal according to the predictions of these four models switching between Persian adjectives and Swedish modifying noun or vice versa should be impossible. The data ruled out such restriction. I found many cases of switching between adjectives and nouns as in (10&11):

10) In *qahve* ***stark**-e*
this coffee strong-is
" This coffee is strong "

11) un ***tröja*** *zard-e*
that sweater yellow-is
" That sweater is yellow "

In (10) a Swedish adjective is postposed to its Persian modyfying noun and in (11) a Swedish noun is switched before a Persian adjective. The two resulting structures are well-formed and satisfy the Persian syntax.

## 8. Matrix language approach: MLF model's system morpheme principle

System morpheme predictions in the MLF model (Myers-Scotton 1993a) were well compatible with the Swedish/Persian codeswitching data. As the system morpheme principle predicts, in my Persian/Swedish codeswitching data, verbal inflections, copulas, help verbs, clitic object pronoun(the suffixes a_/at), e-ezafè( connecting the head nouns and their referents), INFL and the majority of prepositions were realized in Persian. For instance in (12) below we assign the matrix language role to Persian because the grammatical frame of the codeswitched sentence is Persian. Then we find that according to the system morpheme principle, the clitic pronoun, the modal verb, the help verb and verbal inflections are realized in Persian. Whereas the nominal subject and the infinite verb expressing the semantic content come from Swedish:

12) ***Ledaren***-sun mi- xâd ***avgå*** be-kon-e
leader-clitic pro IMPF-want to leave subj-do-3sg present
' Their leader wants to quit '

There were numerous number of above type of codeswitched sentences in data corpus that as the MLF model predicts the system morphemes were realized in Persian.

## 9. Persian compound verbs: a suitable site for insertion of Swedish bare infinitive verbs

In addition to the above possibilities for codeswitching between Persian and Swedish, a suitable site for switching the Swedish items in Persian is the pre-existing frame of Persian compound verb constructions. A very clear candidate for insertion into these frames are Swedish bare infinitive verbs which are pre-posed to Persian help verbs such as *kardan* ' to do '. The combination of Swedish bare infinitve verb and Persian *kardan* creats a Swedish/English version of compound verb construction which is partly identical to monolingual Persian compound verbs.

The difference between the Persian structure of compound verbs and the Swedish/Persian version is that the non-verbal element of Persian compound verb which is either a noun, an adjective or a nominalized form of a verb is replaced by a Swedish bare infinitive verb. The verbal concepts in these constructions are productively expressed by the Swedish bare infinitive verb which has semantic function and the Persian *kardan* ' to do 'which is semantically empty but does bear the verbal inflections. Consider:

14) xodam-o bâ mohit ***anpassa*** kard-am
myself-object with environment to adjust did -1sg past
' I adjusted me to the environment '

This strategy of switching the Swedish infinitive verbs in Persian compound verb frames is very productive and is widely used across the bilingual speakers of the community.

### *9.1 Bilingual speakers establis congruency*

As was sown above, verb compounding is a very productive process in Persian. This productivity is evidenced by the fact that the preverbal element of the compound verbs is not restricted to native Persian elements. There are also Arabic /Perisan compound verbs formed by combination of Arabic nouns which had entered into Persian centuries ago. The following exibits an Arabic/Persian compound verb construction:

***t’lim*** dâdan ‘ to give education ’ ‘ to educate ’

Recent borrowings from European languages also show upp in combination with Persian *kardan*. Compound verbs such as *edit kardan* ‘ to do editing ’, *montaz kardan* ‘ to do installation ’and so on are created by Persian monolinguals in Iran. These mixed types of verb compounding suggest that the pre-planned structure of Persian compound verb construction provide a free frame for receiving the alien elements corresponding to its internal structure. This pattern of Swedish/Persian verb compounding also suggests that speakers are conscious to syntactic incongruity. If they find that elements of one of the languages fits to certain positions in the other, and mixed structures do not violate grammatical principles of either languages, they choose the congruent syntactic plan and codeswitching occurs. In the case of incongruity either they use strategies to establish congruency between incongruent features or avoid to mix these features. Remarkable evidence of a neutralization strategy established by the Persian/Swedish bilingual speakers is the productivity of Swedish/-Persian verb compounding. In this way and by using this strategy the speakers neutralize the clashes which could arise between Swedish and Persian's incongruent verbal inflections. Implicit in this notion of congruity established by the speakers is the idea that limits placed by the grammars block the codeswitching. Bilingual speakers predict these limits and avoid to mixed the codes if they cannot establish congruity.

As illustrated in (15) when speakers find switching between Persian subject pronoun and Swedish VP due to the incongruent INFL features of Persian and Swedish is problematic, and they cannot establish congruity at this switch point, they avoid to mix the incongruent IPs by using a subject pronoun for each clause. Therefore (15) below is a well-formed codeswitched sentence because INFL features are in agreement with the subject of each respective language. A Persian counterpart to (15) does not require pronoun retention in the complement clause:

15) Mâ ke umad-im sued, ***vi har fått flera möjligheter***
we who came-1pl Sweden we have got more possibilities
' We in Sweden have got more possibilities '

As Myers-Scotton(1993) puts it ' using the bare forms may be a strategy to avoid clashes between matrix and embedded language morpheme order. Jakobson proposes that a language accepts only those foriegn structural elements that correspond to its own internal tendencies. Thomason and Kaufman (1988:54) dealing with ' light to moderate structural interference' propose that.' The features incorporated from one language to the other are more likely those which typologically fit well with corresponding features in the recipient language.

## 10. Conclusions

In this study I have examined some instances of codeswitching between Persian and Swedish which is common in the speech of the young bilingual speakers of Persian communities in Stockholm. I have used these data to examine the predictibility of certain models of constraints on codeswitching which are presented in the literature. Investigations of constraints on codeswitching have suggested that codeswitching is prohibited by different language specific or more general and theoretical constraints. The Persian/Swedish codeswitching data ruled out a number of the restrictions predicted by these constraints.

The data corpus used in this study exhibited examples where phrase structure rules and surface structures of the two languages are not equivalent, yet switching between the constituents from both languages has occurred. Switching between verb and objects on the one hand and between noun and adjectives on the other occurred quite often. These switchings occurred despite that Persian is a head first language and verbs take their complements to the left and Swedish is a head- first language and complements to the verb appear to the right of the verb. Swedish predicate adjectives also occurred at postnominal position in Persian context or Swedish nouns appeard before Persian adjectives, regardless that Persian adjectives are postnominal and Swedish adjectives are pre-nominal. Such findings clearly refute claims that switching is impossible if two languages have different linear orders(Poplack, 1980). I also showed that free morpheme constraint ( Poplack, 1980 ) was ruled out when was tested on Persian/Swedish data. In contrast the subcategorization principle proposed by Bentahila & Davies (1986) is compatible with my data. When Swedish single word or two- word phrases were switched to intrasentential position of Persian matrices, they satisfied the subcategorization rules of Persian. When Swedish multi-word fragments were embedded to Persian as complementary clauses to Persian matrices, they retained the sisterhood relations of Swedish.

Further to the above possiblities of switching between Persian and Swedish, Myers-Scotton's (1993) predictions in the matrix language frame model and particularly the system morpheme principle hold true for the data. As is predicted in the system morpheme principle, clitic pronouns, verbal inflections, -e ezafe, help verbs, modals verbs and prepositions were realized in Persian.

The evidences suggested by the present study also reveals that switching between Persian and Swedish is freely possible at all syntactic boundaries other than between a Swedish verb and Persian INFL, between NEG and its complement and between a pronominal subject and a verb.

I have also showed that codeswitching behavior of fluent bilingual speakers is guided properly by syntactic principles. Competent bilingual speakers through the knlowledge that have acquired about the systems of the two languages, make lexical, syntactic and semantic comparisons between the two systems and select a congruent syntactic plan. As long as they find that certain elements of one of the languages fit to certain positions in the other, switching occurs. Otherwise either they employ appropriate neutralization strategies or avoid to mix the codes. A supporting evidence to this claim is the predominant use of Swedish bare infinitive verbs in Swedish/Persian verb compounding through which the speakers simply match the two verbs and create a syntactic verb phrase functioning as a lexeme. The Swedish bare infinitive verb bears the semantic content and a Persian help verb inflects for tense-aspect, person and number. This switching pattern is an innovation, is productive and strongly suggests that young Persian bilinguals have discovered overlapping grounds between the two languages and use it as a common strategy to switch from Swedish into Persian.

*Department of Linguistics*
*Stockholm University*

**Bibliography**

Belazi , H., Rubin, E., & Toribio, A.J. (1994). ' Code switching and X-bar theory. ' *Linguistic Inquiry,* V 25, 2, 221-234.

Bentahila, A., and Davies, E.D. (1983). ' The Syntax of Arabic-French Codeswitching. *Lingua* 59, 301-330.

Chomsky, N. (1981). *Lectures on Government and Binding. The Pisa lectures*. Dodrecht: Foris.

Di Sciullo, A., Muysken, P., & Singh, R. (1986).' Government and code-mixing. *Linguistics* 22, 1-24.

Joshi, A. (1985). Processing of sentences with intrasentential code-switching. In Dowty, D.R., Kartunnen, L.& Zwicky, A. (eds.), *Natural language Parsing*. Cambridge: C.U.P.

Mohammad, J. and Karimi, S. (1992). Light verbs are taking over: Complex verbs in Persian, proceedings of WECOL 1992. CLS 33, 1997.

Muysken, P. (1995). Code-switching and Grammatical theory. In Milroy, L. and Muysken, P. (eds), *One speaker, two languages:*

*Cross-disciplinary perspective on code-sitching*. Cambridge: Cambridge University Press, 177-198.

Myers-Scotton, C. (1993a). *Social motivations for codeswitching*: Evidence from Africa Oxford: Clarendon press.

Myers-Scotton, C. (1993b). *Duelling Languages Grammatical Structure in Codeswitching*. Oxford: Clarendon Press.

Myers-Scotton, C. and Jake, J.L. (1995). Matching Lemmas in a bilingual language competence and production model: Evidence from intrasentential codeswitching. Linguistics 33, 981-1024.

Nishimura, M. (1986). Intrasentential code-switching: The case of language assignment. In Vaid, J. (ed), *Language processing in bilinguals: Psycholinguistics and neuropsychological perspectives*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 123-143.

Nishimura, M. (1997). *Japanese / English Code-Switching Syntax and Pragmatics*. USA: P. Lang.

Pfaff, C.W.(1979). Constraints on language mixing: Intrasentential codeswitching and borrowing in Spanish English. Language 55.2, 291-318.

Poplack, S. (1980). " Sometimes I'll start a sentence in Spanish *y termino en espanol* ": Towards a typology of codeswitching. Linguistics 18.7-8, 581-618

Poplack, S. (1981) Syntactic structure and social function of codeswitching. In Duran, R.P.(ed.), *Latino language and communication behavior.*

Woolford, E. (1983). ' Bilingual Code-switching and syntactic theory.' *Linguistic Inquiry* , V 14, 3, 520-535.

Thomason, S.G., and Kaufman, T. (1988). Language Contact, creolization, and Genetic Linguistics. Berkeley: University of California Press.

# En sprogundersøgelse af en såkaldt halvsproget person

*Frederikke Trondhjem*

## Indledning

Undersøgelsen om halvsprogethed har jeg lavet som en eksamensopgave i sociolingvistik på Ilisimatusarfik Grønlands universitet i Grønland.
Ordet halvsproget er i sig selv et negativt klingende ord, hvor man kan få associationer om en ikke hel person, som ikke har et ikke helt sprog.
Opfattelsen af at være halv grønlænder og halv dansker er beskrevet af Bo Wagner Sørensen, 1997 i 'Når kulturen går i kroppen'. Her beskriver Bo Wagner Sørensen metaforen, om at betragte sig selv delt i to halvdele med hver sin kultur, den danske og den grønlandske. Denne splittelse giver problemer omkring identitet, især ved manglende grønlandsk sprog. På samme måde er halvsprogethed opfattet som det ikke at have et helt sprog, som om det også er en beholder man skal fylde op for at der er et helt sprog.

Da undersøgelsen kun omfatter en enkelt person, er det klart ikke nogen generalisering for dansk sproget i Nuuk, men Rita, (et navn jeg har opdigtet) er en af borgerne i Nuuk, og kan give et billede af sproget. Min undersøgelse er dog i tråd med at undersøgelse af enkelte eller få individer har fået indpas i sociolinguistik og andre fagområder.

## Karakteristik af Rita

Rita er en pige på 21 år, opvokset i Nuuk, og har stort set boet der hele sit liv, undtagen et halvt års ophold i Danmark, da hun var ca. 15 år. Rita har en grønlandsk mor som hun altid har boet hos indtil hun flyttede hjemmefra, og en dansk far der flyttede da Rita var 4-5år, og som nu er grønlandsk gift. Rita taler primært dansk, men kan noget grønlandsk. Hendes mor taler mest grønlandsk til hende men hun svarer for det meste på dansk. Hun bor sammen med en anden blanding (en med grønlandsk mor og dansk far), som primært taler dansk, men sammen forsøger de at få gang i det grønlandske sprog. Rita har en meget karakteristisk piget stemme, som får hende til at lyde barnlig især med det sprog hun har.

Ved undersøgelsen af sproget viser det sig at det barnlige især kommer af, at hun bruger simple sætninger, og hun er vag i sine beskrivelser, dvs. hun bruger ikke så mange beskrivende adjektiver, og hun synes at have et begrænset ordforråd. Det (formodede) begrænsede ordforråd kom til udtryk ved at hun, ved fortælling af dagligdagssituationer, ikke leder efter udtryk og kan tale om det i det uendelige, mens hun ved fortællingen om mere specifikke emner f.eks. dansk og grønlandsk sprogbrug direkte leder efter udtryk der kan dække det hun vil sige.

Rita bruger næsten ingen flerledede sætninger, og hun citerer hellere ordret, hvad hun har tænkt eller hvad andre har sagt, end at sige en mere kompliceret sætning.

En af de karakteristiske træk for Rita er, at hun siger ordene fuldt ud, hvilket jeg også har bemærket hos andre fra Nuuk. Når Rita siger 'lærere' er det også med den sidste stavelse, og 'et valgfrit fag' udtales meget tydeligt.

## Kriterier for halvsprogethed i Grønland

Jeg vil nu gå over til hvordan man afgør om en person er halvsproget. Per Langgård (1992) beskriver følgende miljø, hvor et barn kan blive halvsproget:

*Man skal være en blanding dvs. barn af en dansker og en grønlænder.

*Moderen er grønlænder, som taler et mangelfuldt dansk og har forkert dansk syntaks, forkert intonation og forkert grammatik. Hun har manglende semantisk forståelse og derfor forkert brug af sproget.
*Faderen er dansk, som regel håndværker og næsten aldrig hjemme. Han har ingen eller kortere uddannelse.
*Der tales kun dansk i hjemmet.
*Der er et traditionelt kønsrollemønster, hvor moderen er hjemmegående og står for børneopdragelsen.

Per Langgård mener at den halvsprogede har klare sproglige mangler, der tolkes som tydelige grønlandske træk.
Der er

- manglende stød.
- reduceret aspiration
- reduktion af syntaks (dvs. der er korte sætninger)
- reduktion af leksikon (dvs. begrænset ordforråd)
- overvældende brug af fyldord.

Imidlertid kan det konstateres ved undersøgelse af Rita's sprog med hensyn til disse punkter, at manglende stød dårligt kan bruges som parameter, fordi ikke alle dansktalende har stød dvs. har forskel i f. eks. hund og hun. Man kan heller ikke umiddelbart høre reduceret aspiration. Når Rita siger 'kæreste', er det med tydelig aspiration, og der er distinktion mellem p og b som i 'pjækked', og 'blok'. Da [t] foran [i] udtales ligesom på grønlandsk kan man ikke bruge dette kriterium, selvom PL påstod at der var svagere udtale ved f.eks. tivoli, hvilket jeg er i tvivl om.

Rita har korte sætninger, men grammatisk er de rigtige. Der ingen forkert brug af tid, tværtimod retter Rita automatisk, hvis hun er opmærksom på, at hun har sagt et verbum med en forkert tidsbøjning. Hun har både de regelmæssige og uregelsmæssige verbale bøjninger rigtigt.

Kønsbøjning er heller ikke god som parameter, da der også her er dansktalende uden distinktion i køn. Dog har Rita de rigtige bøjninger som f. eks. 'et valgfrit fag'.

Det sidste punkt 'overvældende brug af fyldord' er til gengæld opfyldt. Eks.:

> så var jeg så væk i p tror det var 14 dage altså ffor mig selv [*hmm*] og så kom jeg så (sukker) over til min moster som bor på Niels Hammekensvej meget p ehh spiste aftensmad derovre [*hmm*] kom og så fjernsyn altså for ikke at være alene *ja* og p jeg boede selv på nummer sytten [*hmm*] og pp så en dag så vil hun så i byen [*fnis*] og det sagde jeg ja til pp så var vi på Kristinemut og så mødte jeg Steen [*ja*] og så var jeg så over og hilse på ham og så [*hja*] hjem og feste og ..
> [Kursiv]= interviewer.

I dette stykke er '*så*' brugt 12 gange som fyldord. I det hele taget bygger Rita sine sætninger mest i paratakser med konjunktionerne *altså, så, og, også*. Men denne parameter kan heller ikke bruges som argument for halvsprogethed, da mange danskere også bruger disse i deres almindelige tale, i specielle sociolekter. Merete Stistrup Jensen (1987) skriver i Kvindesprog: "Både sideordnet sætningsopbygning, brug af tankestreger og prikker, opstoppersætninger og brug af det forstærkende "så" er talesprogstræk" for kvinder. Talesprogstrækkene bruges til at markere, at manden har højere status end kvinder, hvor kvinden ikke vil fremstå som mandig, men menes også at øge tryghed, sprede velbefindende hos familie og venner.

Konklusionen er at Rita har et grammatisk korrekt sprog, som har sit særpræg ved at have forholdsvis korte sætninger, begrænset ordforråd og en meget tydelig udtale.

## Halvsprogethedsteoriens afstamning

Halvsprogethedsteorien som Per Langgård tager udgangspunkt i stammer fra Nils-Erik Hansegårds 2 bøger Tvåspråkighet eller halvspråkighet? 1968 og Den Norrbottensfinska Språkfrågan, En återblick på halvspråkighets debatten, 1990.

Hansegård opstiller kriterier for halvsprogethed som består af 6 punkter hvoraf de første tre er kvantitave og drejer sig om ordforrådsstørrelse og sproglige elementer, sprogrigtighed og graden af automatik, mens de 3 sidste er kvalitative og ret svære at vurdere. Det drejer sig om evnen til at skabe og nyskabe, beherskelse af sprogets tanke-, følelses-, og viljefunktioner og rigdom eller fattigdom af betydningsoplevelser.

Forskere som Wande har prøvet at efterprøve de kriterier Hansegård sætter for halvsprogethed hos Norrbottens finner, men jeg kender ikke resultatet. Der har været meget kritik af selve begrebet halvsprogethed, (semilingualisme) og nogle forskere som Skutnab-Kangas mener at det er et politisk våben. Således bruges begrebet også i Grønland, da Per Langgård selv bruger det som en kritik af skolepolitikken i Grønland, hvor danskfaget i folkeskolen har fået færre timer efter Hjemmestyrets indførelse.

Jeg vil ikke gå ind i detaljer om denne debat, men tilslutte mig Suzanne Romaine's syn på sprog i *Bilingualism* 1995 i afsnit 6.4, at man videnskabeligt ikke kan fastsætte rammer for sproglig kompetence, men at der er forskellige faktorer, der spiller ind i sprogtilegnelsen blandt andet personlighed, input fra omgivelserne og cognitiv udvikling.

## Input fra omgivelserne

I det følgende vil jeg kigge på input fra omgivelserne, der kan påvirke Rita's sprogtilegnelse.

Ser man på sprogsituationen i Grønland, er den de jure monolingualt grønlandsksproget og de facto bilingualt i grønlandsk og dansk. Dansk har været prestigesproget siden daniseringsperioden som startede i 50'erne. I optakten til hjemmestyret vendte billedet, hvor grønlandsk fik større betydning, og skolepolitikken gik ind for mere grønlandsk undervisning og mindre dansk undervisning. Og her kommer så de monolinguale dansksprogede blandinger i klemme, fordi deres manglende danskfærdighed ikke opfølges i skolen. I en pilot projekt sammenlignede Per Langgård eksamenskaraktererne i faget dansk hos de rent dansksprogede blandinger og aktivt eller passivt bilinguale. Undersøgelsen viste at de monolinguale blandinger havde et lavere gennemsnit end de bilinguale. Denne forskel blev tolket som et udtryk for halvsprogethed. Det gav bl.a. identitetsproblemer i form af ikke at kunne opnå sociale attraktive miljøer og uddannelsestilbud pga. manglende grønlandsk sprog samtidig med at være en 'mangelfuld' dansker rent sprogligt. Dvs. man kan hverken få de lavere uddannelser som kræver grønlandsk sprog, eller de højere uddannelser der kræver dansk sprog.

En alternativ tolkning af dette resultat er at kigge på den sproglige lagdeling i samfundet. Vi ved at den grønlandske befolkning enten kun er grønlandsktalende eller grønlandsk-/dansktalende. Den danske befolkning er mest dansktalende med få undtagelser der kan grønlandsk. Foruden disse er der de blandede mono- og bilinguale.

Dansk er stadig prestigesproget, da alle de høje stillinger forudsætter at man kan tale dansk, og bilinguale har højeste forudsætninger for at få arbejde. Den grønlandsktalende befolkning siges ikke at have sproglig lagdeling, men afviger fra ‘standardsproget’.

## Kammeratskabet og skolen

Betydningen af kammeratskabet, som understreget af Labov, har ikke i så høj grad haft betydning for Rita’s sprog. De kammerater Rita har nævnt er kun grønlandske med dansk som L2. Hendes barndomsveninde kunne ikke så meget dansk i starten, men lærte det efterhånden. Hendes tidligere kærester i teenage alderen er grønlandske. Rita lærte således ikke mere dansk af sine kammerater, snarere tværtimod var det hende der kunne dansk, og kammeraterne lærte snarere af hende. Hun gik i en dansk klasse, fra 2.klasse, og fik lært de grundliggende grammatiske regler, de forskellige bøjninger, kønnet, og ordstillinger på den rigtige måde.

Hun havde grønlandsk som fremmedsprog. Hun mener ikke at have lært så meget grønlandsk, da klassekammeraterne tit var nye fra Danmark, og klassen skulle starte forfra hele tiden. Hun har gået i skole indtil 10. Klasse, som er svarende til dansk 9 klasse, og går nu på handelskolen, hvor hun går i en grønlandsk klasse.

Rita startede sin skolegang ca. 1980, omkring Hjemmestyrets indførelse. Antallet af dansksprogede børn toppede omkring 1970-80. Attituderne blandt lærere i skolerne og folk i almindelighed var, at der var for mange dansktalende børn, i hvertfald i Nuuk, og derfor måtte man gøre en indsats for grønlandsk, hvis man skulle bevare sproget. Der blev ikke gjort så meget for at udvikle danskundervisningen, men til gengæld styrkedes grønlandsk.
Jeg tror ikke blot at det er danskundervisningens kvalitet der har resulteret i Rita’s sprog. Ritas skolegang endte i 10. klasse, hvor hun netop var igang med at lære sit voksne sprog.

Det er muligt at Rita’s sprog havde udviklet sig mere, hvis hun havde dette er ikke yderligere undersøgt. Den dansktalende befolkning har ikke været undersøgt før i år af Birgitte Jacobsen fra Grønlands Univesitet.

Til undersøgelsen af den sproglige lagdeling har jeg ladet mig inspirere af den kontroversielle kodeteori af Basil Bernstein (1974). Bernstein opdeler sproget i restringerede koder, som bruges i forbindelse med arbejderklassen og elaborerede koder i forbindelse med middelsklassen. I middelklassen opdrages børnene i universalistiske betydningsindhold, dvs. hvor børnene opdrages mere kontekstuafhængigt, og børnene lærer at tale eksplicit til deres omgivelser, mens der i arbejderklassen opdrages i partikularistiske betydningsindhold, dvs. børnene opdrages mere kontekstafhængige og lærer at tale implicit til deres omgivelser. Bernsteins tese er at modtageligheden for uddannelse har en forbindelse med socialstruktur og symbolstruktur, hvor skolen bruger de elaborerede koder med det resultat, at børn med restringerede koder ikke klarer sig så godt i skolen.

Teorien har været meget kritiseret bl. andet af Labov (1984). I sin artikel om undersøgelse af Non-standard Negro English, siger han at sorte skolebørn fra lowerclass og workingclass taler fint i forhold til hvad de har lært i hjemmet og på gaden. Han mener at den sociale situation betyder meget for barnets verbale udfoldelse og hvis der ikke er tryghed og balance i den kommunikative situation vil barnet tie. Skolesituationen er en assymetrisk kommunikativ situation, hvilket er ugunstigt for en 'mærket' elev, og det er som regel de gode elever der får skulderklap.

Kammeratskabets betydning beskrives hos Labov som den faktor der vedligeholder børnenes sprog, hvor de skal have et specielt sprog for at være accepteret indenfor kammeratskabet. Dvs. at børnene har en bestemt sociolect der afviger fra standard engelsk og som af lærerne bliver opfattet som et forarmet sprog.

Både Bernstein og Labov er enige om at arbejderklassen har et anderledes sprog, og lærerne i skolen ikke formår at afkode børnenes sprog, fordi de fortsat i skolen, da det ser ud til at hun mangler input for at få lidt mere nuancering i sit danske sprog. Men det må ligeså vel være det miljø hun er i, der holder hendes sprog "ved lige" dvs. hun kan ikke så godt kan ændre sit sprog af sociale grunde.

## Rita's sociale netværk

Ser man på Rita's sociale situation har hun en grønlandsktalende mor, uden fast arbejde, mens faderen har et godt vognmandsfirma. Faderen er en af de danske håndværkere og arbejdsmænd der kom til Nuuk i løbet af 60'erne og 70'erne, for at udføre et arbejde, og hvor nogle af dem har bosat sig og giftet sig grønlandsk. Mange håndværkere startede firmaer og blev rige, og dannede derved en overklasse, der ikke rigtig eksisterede før de kom. Før dem var det de overordnede i Kongelig Grønlandske Handel, skolevæsenet, kirkevæsenet og de få uddannede grønlandske lærere, der dannede en overklasse, men der var ikke rige folk, som grundlag for det borgerlige liv, som man ser i Europa.

Overklassen blev efterhånden håndværkerne og de handlende, der ikke havde et sprog af borgerlig oprindelse fra Danmark. Mange af håndværkerne kom fra Jylland, og det statusgivende sprog blev en blanding af disse menneskers sprog. Mange af håndværkerne var gift med grønlandske kvinder, der ikke var perfekt dansksprogede. Disse mennesker holdt meget sammen og blandede sig ikke med de korttidsansatte danskere. Efterhånden dannedes der et sprog der udvikledes til et statussprog specielt i Nuuk, som afviger fra det danske standardsprog. Det er tænkeligt at mange danskere der har været i Grønland i mere end 25 år efterhånden har ændret deres sprog til det sprog, der tales i omgangskredsene.

Alt dette har som sagt ikke været undersøgt og må tages som min og mine medstuderendes erfaringer, men giver grund til at tro, at der er et specielt dansk sprog i Nuuk og måske også andre steder.

## Det social psykologiske aspekt hos de bilinguale

Til slut vil jeg kigge på det social psykologiske aspekt hos de bilinguale.

Lamberts matched-guise undersøgelse af fransk-canadiere og engelsk-canadiere i Canada har blandt andet vist, at teenagepiger har et tættere psykologisk forhold til deres far, idet de favoriserer faderen fremfor moderen, som krydsidentifikation, men at denne favorisering bliver mindre i midten af teenageårene eller senere. I Ritas tilfælde er faderen tilfældigvis dansker, men ovenstående kan forklare hendes valg af sprog, trods det at hun bor hos sin grønlandsk talende mor. Lambert skriver videre at de etnocentriske tendenser og attituder mod den anden gruppe hos den der lærer et andet sprog (L2) er bestemmende for motivationen af indlæringen af L2. Motivationen bestemmes enten som instrumental, hvis det er nytten af sproget der er målet i beskæftigelse, og integrativ hvis det er for at blive part af den anden gruppe. For Rita er grønlandsk en del af hendes identitet, som hun fastholder især omkring højtidstraditionerne, og det lyder som om, at hun har fordelt de to kulturer med dansk som instrument til at komme videre i uddannelsen med, (samtidig med at hun er intergreret i den danske kultur), men har valgt at holde på de grønlandske traditioner, som hun synes er mere "menneskelige".

Prestigesprogets indflydelse er også undersøgt af Lambert i ovennævnte matched-guise undersøgelse, som viste, at de mennesker der ikke behersker prestigesproget som modersmål, føler deres sprog som "mindre værd" og denne tendens er også at finde i det grønlandske samfund. Attituderne til det danske sprogs mere instrumentelle værdi betyder selvfølgelig også, at en blanding vælger et prestigesprog som sit primær sprog.

## Konklusion

Som konklusion mener jeg at kriterierne for halvsprogethed som Per Langgård skriver om, efter hvad jeg kan bedømme ikke kan bruges til bestemmelse af halvsprogethed. Det er snarere et spørgsmål om begrebet halvsprogethed giver mening i denne sammenhæng. Ud fra de undersøgte faktorer mener jeg der snarere er en særlig dansk dialekt i Nuuk, som Birgitte Jacobsen er ved at undersøge.

## Litteraturliste

Bernstein, Basil: Sprog, socal klasse og socialisation. In Gregersen, Frans m.fl. Klassesprog. Sociolinguistik og uddannelse - en antologi. Borgen Basis, København 1974.
Corder, S. Pit: Sprogundervisning og sprogvidenskab. Gyldendal 1977.
Hansegård, Nils-Eril: Tvåspråkighet eller halvspråkighet? Bokförlaget Aldua/Bonners. Stockholm 1968.
Den Norrbotttensfinska Språkfrågan. En återblick på halvspråkighets-debatten. Uppsala Multiethnic Papers 19. Centre for Multiethnic Research. Uppsala University, Uppsala 1990.
Hansen, Erik: Sprogiagttagelse. Gjellerup 1973.
Hansen, Erik og Lund, Jørn: Sproget her og nu. Dansk Sprognævns skrifter 14. Gyldendal 1988.
Jensen, Merete Stigstrup: Kvindesprog. Ålborg Universitetscenter 1987.
Jensen, Per Anker: Principper for grammatisk analyse. Nyt Nordisk Forlag Arnold Busck, Københaven 1985.
Joseph, John Earl: Eloquence on Power. Francis Pinter, London 1987.
Labov, William: Socilinguistic Patterns. University of Pensylvania Press, inc. 1984
Lambert, Wallace E: A Social Psychology af Biligualism. Journal of Social Issues Volume XXIII, Nr. 2, 1967.
Langgård, Per:Grønlandsksproget - Tosproget - Grønlandsksproget.: Nogle tendenser i det dansk-gørnlandske sprogmøde blandt Nuuks skolebørn. In: Grønlandsk Kultur - og Samfunds Forskning 92, Ilisimatusarfik/Atuakkiorfik, Nuuk 1992.

Milroy, Lesley. Observing &Analysing Natural Language. Blackwell Publishers, Oxford 1993.
Romaine, Suzanne: Bilingualism. Basil Blackwell 1995.
Teleman, Ulf: Ret og rimeligt. København, Hans Reitzel 1981.
Øzerk, Kamil Z.: Om Tospråklig Utvikling. En teoretisk studie av fenomenet og begrebet tospråklighet. Oris forlag, Oslo 1993.

# ENDANGERED LANGUAGES

# How fast will a language die when it is officially no longer spoken?

*Ida Elisabeth Mørch*

## Abstract

This paper reports on field work on dialects of Kalashamon, an Indo-Aryan language spoken in NW Pakistan. The people concerned are Muslims who have converted from the Kalasha beliefs to Islam recently and subsequently shifted language from Kalashamon to Khowar, the language of the dominating Kho-people. It is found that Kalashamon has not been given up as early as claimed. Especially among the women the language is still used, probably due to a strict separation of the male and female spheres of lives which gives room for two language codes in the society: a domestic one and a public one.

## Introduction

In this paper I will discuss some of the factors that have influenced a process of language death. I will base my discussion on studies of four dialects of Kalashamon, an Indo-Aryan language spoken in North West Pakistan.[1] According to local information these dialects have not been spoken for three or more generations due to a shift of language and religion. This situation can be characterized as 'radical language death' (Campbell & Muntzel 1989: 183). I will question how radical this language death has been and I will concentrate on the ethnographic circumstances of the shift from one language to the other. I will show that well hidden behind religious and cultural barriers a language can still be spoken. This will give an idea of what can keep a language alive in spite of a strong social pressure. And it will illustrate some of the methodological problems that we faced in this difficultly accesible area.[2]

### The Kalasha people

The Kalasha number about 3,000 people who live in three narrow side valleys to the larger Kunar Valley in the Chitral District in the western part of the Hindu Kush range in Pakistan, close to the border to Nuristan in Afghanistan.

The Kalasha are the only non-Muslim people in the multicultural and multilingual Chitral District. They practice a religion which anthropologists characterize as pre-Islamic with ancient vedic traits. Until the beginning of this century the Kalasha settlements formed part of a larger area known as Kafiristan, Land of the Unbelievers. For centuries the Kalasha have been serfs under the ruling (Muslim) Kings of Chitral, the Mehtars and in that way they have taken an inferior position compared to the surrounding Muslim peoples.

Despite official protection from the Pakistani government as an ethnic and religious minority the Kalasha are in the daily life exposed to the very aggressive neighbouring Islamic society, specifically from the largest ethnic and linguistic group, the Kho people. More and more Kho people have moved into the Kalasha valleys were they have bought land and live as farmers and shopkeepers, and today every Kalasha village has its Muslim twin with a Muslim run school, shop and a mosque. And every week the Kalasha see family members convert to Islam. With the ungoing high speed of conversion and the following language shift to Khowar, the language of the Khos and the lingua franca of Chitral, the Kalasha should be regarded as a threatened culture and Kalashamon as a threatened and probably dying language, although this last process could be slower than the religious shift.

## Former Kalasha settlements

We know from the oral literature, local history and several reports from the beginning of this century that Kalasha settlements were to be found several places in southern Chitral. In these places all inhabitants are now Muslim, and the population is a mixture of Kalasha descendants, Kho-people, Pathans and descendants of Nuristani immigrants.

South of the present Kalasha area there are two side valleys where conversion from Kalasha to Islam presumably took place between 1920 and 1950. Many of the Kalasha descendants here still speak Kalashamon in the home, and even the children learn the language together with Khowar.

Besides these five valleys where Kalashamon is still spoken there are a number of villages in the main valley and in an eastward side valley where Kalashamon was spoken until the 1940'ies-1950'ies or later. Today the inhabitants of these villages speak Khowar (except in Kalkatak where they for historical reasons speak another Dardic language called Palula). These villages (Suwir, Gromel, Kalkatak, Lawi, and Birga) where Kalashamon is officially no longer spoken are the focus of this paper.

## Problems linguists meet in the investigation

The investigation of the retainment of Kalashamon requires caution and tact. Generally there is a very contemptuous attitude towards the use of Kalashamon in the area because of the low status associated with the Kalasha.

Furthermore, there is a culturally determined separation of the male and female spheres of life in this area. It was therefore a great advantage for the linguistic investigation in this culture to have a research team with both a man and a woman. We had the idea that Kalashamon was retained better in the female i.e. non-public world than in the male and public, so because I had the possibility to visit the women in their home there should be a chance of getting near the former or maybe even present speakers of Kalashamon. A problematic factor, however, was that our local contact-person would typically be a higher-ranking man of Kho origin, because in this culture you need a separate guest-room or a separate inner courtyard to house strangers properly - even if it is only for an afternoon. The people of Kalasha origin had in most cases retained a lower social position and would not be considered appropriate to house strangers.

I was usually at some point invited inside the house to meet the women. But the Kho origin of the family meant that I did not have the great opportunity to investigate the Kalasha language as we thought. Another limitation on my part was that I do not speak either Khowar, Pashto or Urdu. This means that I had problems of just asking about the knowledge of Kalashamon and to ask if the women could take me to someone who spoke the language. Sometimes I had a male accompaniment from the family to interpret (if someone knew English) but then as long as he was present no one would admit to know Kalashamon.

## The findings

In Kalkatak the speakers of Kalashamon had made a decision of shifting from their old language to Palula about 20 years ago, so if this had in fact happened we would have a 'radical language death' situation. Our male informant at about 60 had not used the Kalasha language for a long time. But after a while he remembered many Kalashamon words, and it was clearly, as found by Campbell

& Muntzel (1989: 184), the more frequent and salient lexical items that were remembered first.

In Kalkatak I worked with two ladies aged about 50 who still spoke Kalashamon fluently and, contrary to what the men said, still used it among themselves.

I also worked with the wife of our male informant. She was about 55-60 years old and remembered the words on my word list quite well. We were sitting in the kitchen surrounded by her daugther and daugthers-in-law, and sometimes they contributed with a word if the old lady did not remember it. They were about 25-35 years old and must have heard the language spoken in their childhood. I think that they had never been fluent in Kalashamon but should be characterized as 'rememberer' (Campbell & Muntzel 1989: 181).

In Suwir, a village opposite Kalkatak in the main valley, we were told by different people that between 10 and 100 persons still remember the language. Here we had opportunity to interview two elderly men and two elderly women. They had spoken Kalashamon when they were young and can be caracterized as 'once-fully competent' speakers (Campbell & Muntzel 1989: 184). They admitted to have used the old language for a long time as a secret language to keep secrets for their children and for chatting among friends. The men were capable of having a conversation in Kalashamon. Without understanding all of it we hear many fixed phrases and a limited use of inflections. The women did not allow us to tape record the interview.

Also in Suwir we at one point went from one house to another. I was walking behind the others when a woman looked out from a house and addressed me in Khowar. I did not understand what she said and told her that in Kalashamon after which she *in Kalashamon* said to me that she did not know Kalashamon!

The village of Lawi has about 600 inhabitants and about two thirds of these are of Kalasha origin, the rest are Kho-people. It is our impression that the knowledge of Kalashamon is relatively widespread in this village but it was very difficult to obtain any data here because of a very strong aversion to Kalashamon.

In Lawi we spoke with a male informant aged 35 who could speak the language. It was claimed he was the youngest person in the village who knew Kalashamon. He said that he had learned it from his mother because he wanted to. Another man in Lawi was our informant for only a short time, then he was called home because his wife did not appreciate that he was speaking the old language in the village square in public.

In Gromel, a part of Drosh, the attitude to the past and to Kalashamon was also very negative. So no-one volunteered to work with us and we think that the woman who finally turned up was forced to by our host. She had originally come from another village, and presumably had a very low status in the village. She tried very hard to help us with words in her old mother tongue but very often she provided us with Khowar words instead.

These examples illustrate why we think that the language has been spoken for a longer time than people at first claim. Our findings about the time of conversion are thus in agreement with those of the anthropologists Cacòpardo and Cacòpardo as expressed in three papers from 1991.

From my visits to the women in the houses we got the impression confirmed that the Kalasha language has been retained even longer in the domestic and non-public sphere than in the public life.

## Reasons for the retainment of Kalashamon

The main reason for this retainment more or less hidden from the public is to be found in a culturally and religious clear-cut separation of male and female spheres.

In this area according to both the Islamic and the cultural tradition the male world is associated with the public life. Men go to the bazaar, men go to the prayers in the mosque, men are chatting in the village square etc. The female sphere is primarily associated with the home and takes place mainly inside the house. The women meet each other during the day in their homes and in the fields. The restriction laid on their freedom of movement is that they cannot go outside in the public without veil and male accompaniment. The restriction on where a man can go is into other people's house except if he is closely related to one of the women in the house.

This strict separation of the spheres of men and women has given room for the women using Kalashamon in their daily life among themselves as a sort of domestic language code. The users of this language code are less exposed to the cultural and religious pressure in the public life where it is important to show oneself as a devoted Muslim and dissociate oneself from the pagan past. Correspondingly it is natural for the men who go about in the public to speak the language of the Islamic surroundings, namely Khowar.

There are some other reasons for a retainment of Kalashamon for some time after conversion.

One is the geographical isolation in the very narrow side-valleys where the powerful Kho-people did not come often. The way that trade has taken place in former times was that non-local travelling tradesmen came into the valleys, and the men in the village would from time to time go to buy necessities in the main valley. This may have delayed the language shift in the valleys Jinjiret, Urtsun and Shishi Kuh.

Another factor is the historical connection between the various former Kalasha communities due to patterns of intermarriage. The common low social status of all Kalasha may also have reinforced a strong interconnectedness between these communities and delayed the language-shift. This factor has relevance especially to the villages in the main valley, Suwir, Kalkatak, Gromel and Lawi.

I assume that these factors together have kept the Kalashamon dialects alive for a longer time than has officially been claimed.

## Conclusion

Due to the difficulties of the investigation of these dialects our material is not representative enough for making specific claims about the linguistic consequences of the language shift. Still we judge the phonology of the former language to be fairly intact and have identified some special features (not present in the dominant language) being retained in the dying language, quite in accordance with Campbell & Muntzel (1989: 183). In addition we have seen grammatical productions characterized by fairly simple constructions and phrases, as Campbell & Muntzel mention for one of their once-fully-competent speakers (Campbell & Muntzel 1989: 184). Furthermore, some of the dialects on certain points seem even more archaic than the dialects still spoken in the western valleys which also make these dialects interesting.

An answer to the title of my paper "How fast will a language die when it is officially no longer spoken?" can then be that it is very hard to know. Because the speakers themselves do not know if someone in the village still speak the language. And they are unwilling to admit it if they do know the language. Our scientific wish to reveal the hidden use of Kalashamon then clashes with the speakers rights to decide 'what should or should not be done with their language' (England 1992: 33) i.e. if the language should be maintained or abandoned. Moreover a religious barrier and the conflict between western culture versus the local rural culture also prevents clear insight for the investigator.

We found remnants of Kalashamon dialects that would be very interesting to study, much more for historical, linguistic and ethnographic reasons. Unfortunately they are very difficult to get to and from what we know they are dying at a high rate.

*Ægirsgade 3*
*DK-2200 Copenhagen N*

**Notes**

[1] Kalashamon is one of the little known and studied so-called Dardic languages of the Indo-Aryan branch of the Indo-European languages. The dialects concerned are important for historical phonological reasons and also for language contact reasons. It is related to but unintelligible with Khowar.

[2] The field work was carried out together with Jan Heegaard in three periods from 1995 to 1997. It was conducted with financial support of His Royal Highness Crown Prince Frederik's fund, The Danish Research Council of the Humanities, The Faculty of the Humanities of University of Copenhagen, and stud.mag. Per Slomann's Foundation.

## References

Cacòpardo, Alberto & Augusto Cacòpardo. 1991. The other Kalasha. A Survey of Kalasha-mun-speaking people in Southern Chitral, Part 1-3. East and West I-III. 41.273-373. IsMEO.

Campbell, Lyle & M.C. Muntzel. 1989. The structural consequences of language death. (Dorian, Nancy C. ed.) Investigating obsolescence: Studies in language contraction and death. Cambridge: Cambridge University Press. 181-196.

England, Nora C. 1992. Doing Mayan linguistics in Guatemala. Language 68.1.29-35.

# Komi, an endangered language?

*Marja Leinonen*

**Abstract**
Komi, or Komi-Zyrian, is an Uralic language indigenous to the Republic of Komi in the North-Eastern corner of European Russia. It is spoken by some 300,000 inhabitants of the republic and by several thousand outside its borders. In 1992, the language was given official status alongside with Russian. A revitalisation programme began in 1994, and this paper is an interim report on the progress made so far (summer 1998). The main factors affecting the maintenance or loss of a minority language are applied to Komi.

## 1. Introduction

On September 25, 1998, the Parliamentary Assembly of the Council of Europe accepted a resolution on the endangered Uralic minority cultures which, among other things, urged the Russian Federation and other countries where they are present to support these people's languages, cultures and traditions, through education in their own mother tongues and through publishing and mass media in their languages (Doc. 8126, June 1998). Among the Uralic languages, Khanty, Mansi, Nenets, Karelian, Vepsian, Livonian, Votian, Izhorian, and most Samoyed and Sámi languages are endangered to varying degrees. Those languages that, according to the report, 'may be endangered' are Mordvin, Mari, Komi and Udmurt. Komi in this context covers both Komi-Zyrian, spoken in the Komi Republic, and Komi-Permiak, spoken in the Autonomous Komi-Permiak district. Both are literary languages, though they are very close to each other.

Komi-Zyryan is not in immediate danger, rather, it is a language for which there remains some hope: despite ongoing language shift into Russian, it has legislative support and an official programme of revitalisation. Still, as often reported, the economic situation in Russia defines the priorities at present to the detriment of the realisation of the programme. We may be witnessing the last efforts to counteract full Russification of a formally bilingual state. The same can be said for the rest of the Uralic languages spoken in the Russian Federation.However, as the Republic of Komi was the first of the autonomous Soviet socialist republics with a Finno-Ugrian population to recognise the revitalisation potential created by the fall of the Soviet Union, the situation is worthy of an interim report.

## 2. Komi Republic: population

In the north-eastern corner of European Russia, the Komi autonomous area was created in 1921. In 1936 it was reconstituted as the Komi ASSR, and in 1992 as the Republic of Komi, a part of the Russian Federation. In the census of 1989, the total population of the republic was 1,251,000, spread out over approx. 400,000 square kilometres, which thus makes Komi one of the least densely populated regions of Europe. People of Komi nationality reached 291,500, which was 23,3% of the population (nationality was determined in the internal passports, where it depended partly on one's parentage, partly on bureaucratic decisions of unknown motivation). This represented a steady proportional decline since the high of 92% of 1926. Other nationalities in the 1989 census were: Russians (57,7%), Ukrainians (8,3%), Byelorussians (2,2%), Tatars (2,1%) and others (6,4%; altogether over 100 nationalities). Elsewhere in

Russia, people of Komi nationality live on the Kola peninsula and other parts of Arkhangelsk province (over 7,000), in the Yamalo-Nenets area beyond the Ural mountains and elsewhere in Siberia (over 10,000) (Lallukka 1998: 315). However, 87% of the Komi live within the republic. 53% of these live in the countryside. Even in rural areas in several administrative districts and in the towns Russians are in the majority, and Russian is also the urban language of communication (Entsiklopediia. Respublika Komi. Vol. 1, 1995).

By 1994, the total population in the republic had diminished by 5%. The percentage of Komi, due to out-migration of mainly Ukrainians, Byelorussians and other nationalities, rose in 1994 from 23,3% to 26,3%, while the Russian level remained at 57,7%. Another factor was that people were more willing to label themselves as members of the titular nationality. The percentage of those of Komi nationality considering Komi as their mother tongue rose from 70,4% in 1989 to 73,6% in 1994 (note that this was a 'micro-census', only 16,387 were asked). At the same time, it came out that the percentage of people with Komi nationality speaking Komi at home was only 43% (Spravka o natsional´noi situatsii v Respublike Komi).

### 3. History of contact between Komi and Russians

At the end of the XIth century the East Slavic city-state of Novgorod began to collect tribute from the peoples along the rivers of the European North-East. Soon, colonists appeared in the Western and Southern areas inhabited by the Komi. The result was partly assimilation of the Komi, partly their migration further to the north and east. Later, Moscow administered the taxation of the Komi for the Golden Horde, and after the fall of the Kazan Khanate, for itself.

From the XVIIth century onwards, Russian settlements arose along the northern rivers, and later, Russians were settled in the few industrial centres that arose in the area. Before this century, there were, however, no large migrations of Russians to the Komi region. Thus, in 1917-18, within the present borders of Komi, there lived 190,000 Komi and 20,000 Russians (Zherebtsov & al. 1996:52-90, 114).

Though Komi had also been an area of exile for the Russian Empire, it became specialised for this purpose during the Soviet period. In the 1920's, there were established camps with the aim of exploiting the area's natural resources, mainly coal, oil and forests. The first group of prisoners arrived in 1920 in Ukhta, and by 1936 their numbers rose to 100,000. In 1942, there were 200,000 prisoners in Komi - which surpassed the native population of the region (Zherebtsov & al. 1996:220). In fact, Komi had the densest network of prison camps in the entire USSR. At the end of 1950's, when the camps began to be dismantled, the 'spetsposelki' (special settlements) for the exiles were renamed free settlements. All this was planned directly from Moscow to meet the needs of the USSR. The local population was subordinate to the process, serving only the forest industry.

In the 1950's the oil, coal, gas and pulp and paper industries drew workers from outside; often, they also left soon after. The increase of the Russian population has thus been dictated by industrial needs, not by natural population increase. Simultaneously, the Komi population was marginalised both in economic and cultural respects. The urban population (including the ex-camp settlements) in 1994 amounted already to 74,7%. The number of villages went down in the 60-70's due to the All-Soviet programme of 'liquidating villages with no prospects' (Entsiklopediia: Komi respublika, Vol. 1, 1995).

## 4. Language and education

Though Komi is considered to be a young literary language, there was a written language and a special alphabet created already at the end of the XIVth century. Apparently the alphabet was not widely employed, and it was forgotten. The origins of writing came with the Christianisation of the Komi, which subjected them to Moscow. The Cyrillic alphabet was in use at least since the XVIIth century. In the XIXth century, grammars were written and texts translated, there were the beginnings of belles-lettres and folklore collecting, based on different dialects. The language of the schools, such as there were, was Russian (Entsiklopediia. Respublika Komi, Vol. I, 1995).

After 1917, Soviet ideology at first promoted the maintenance of indigenous languages. In the 20's, literacy campaigns, the creation of scripts and the codification of languages were undertaken and employed a great number of linguists. The policy was 'korenizatsiia' (indigenisation), or 'Komi-Zyrianisation' , i.e. educating an indigenous intelligentsia and administrative elite (Popov 1990). Thus, in the 20-30's Komi was in fact an official language of the area. For instance, in 1925/26, out of all primary schools, 203 (75%) worked in Komi, and 54 in both Komi and Russian (Tsypanov 1997). Printing was hampered, though, by changes in the alphabet - in Komi it was altered three times during the 1920-30's.

Since the 30's, Soviet language policy was aimed at raising the Russian language skills of the population and promoting the so-called 'national-Russian bilingualism': education in the native tongue was seen as an important stage which prepared the children for the switch to Russian (Grenoble & Whaley 1998:47). Not accidentally, the 30's were the time of Stalin's purges (repression

420

of 'spies, internationalists and nationalist bourgeoisie') and collectivisation. Linguists, writers, teachers, culturally and politically prominent Komi were sent into exile or shot as nationalists.

In the 40-50's, the Komi language was driven out of administrative use.In 1958, a law on the facultative status of teaching languages allowed parents to decide which would be the school language for their children. The results were that since Russian was needed anyway for social mobility, most parents chose Russian. The language of all official business and higher education was Russian. Consequently, the native language had low prestige.

In the 1960's, many rural schools were closed down, as the policy of closing down unprosperous villages developed. In 1965/66, only 8% of the schools worked in Komi, in 1970/71 the figure went down to 6% (Kratkie svedeniia).

In the 1970's, Komi-Russian bilinguals constituted roughly 70% of the nationality. The rest of the Komi were monolinguals, commanding either Komi or Russian. Over the 1970's, a shift took place: the percentage speaking only Russian grew. This linguistic assimilation took place both within and between generations. The formation of the bilingual layer had begun before World War II, and it continued intensely afterward, especially in towns. This layer largely failed to transmit their mother tongue to their children. In the 1980's, Komi-Russian bilingualism turned into a general characteristic of ordinary Komi, including rural inhabitants. Since World War II, intermarriage has increased, and the second-generation consequence is Russification - the out-marriage rate was about 40% in the 70's. Industrialisation and urbanisation inevitably lead to the same result. Due to these trends, the Finnish demographer Seppo Lallukka envisaged in the late 80's a catastrophe for not only Komi, but for all the East Finnic languages (i.e. Udmurt, Mari, Mordvin) (Lallukka 1990:262-264).

When the perestroika started in 1985, there were no schools with Komi as the teaching language (although in the countryside, Komi teachers had been forced to resort to explaining Russian texts to the pupils in their mother tongue). Now, however, the administrative closing down of 'national' schools (i.e. with Komi as a subject) stopped. Interest in the Komi language began to rise. From 1988 to 1991, the number of 'national' schools rose from 166 to 255 (from 15,000 pupils to 21,700) (Kratkie svedeniia).

## 5. Language legislation and revitalisation

The fall of the Soviet Union and the Communist Party did not affect the situation of Komi immediately, but it was important in reinforcing their ethnic identity. There were even some voices calling for independence, but the main line advanced has been that of moderation. There are several organisations which propagate the rights, and even special rights, of the indigenous Komi (Shtrikhi 1994:23-35).

Post-Soviet nation and state building has reinforced the republic's political identity as a sovereign nation-state within a federal Russia. Political and economic interests have coincided, and the indigenous population's demands for 'national rights' have supported the institutional desire for greater self-government and more economic powers in relation to Moscow (Fryer 1998:583). International agreements on minority and indigenous rights are taken into consideration in the new constitution. For example, there is a law 'On the status of the Congress of the Komi Nation'. This Congress, founded in 1989, can only make proposals for legislation. So far there have been five congresses, and at the moment, the earlier demands for more power in the form

422

of, e.g., a two-tiered parliament with a Komi chamber, have been dropped (Kiselev 1998).

In 1992, the law on the official languages of the Republic of Komi was accepted. It declared these to be Russian and Komi, and Komi as an indigenous language to be the special concern of the state, while Russian was the main language of communication between the nationalities (Strikhi 1994:129-133). A Ministry for Nationalities was founded to put the law into action. Since 1995 it has included an informational translation section, the task of which is to translate all administrative documents into Komi (Acts of the State Council, decisions of the Constitutional Court, Acts by the Ministries, etc.), as dictated by Articles 7 and 10 of the Law. It is also responsible for the Programme of developing and maintaining the Komi language, which became effective and began to be realised in 1994. Naturally, the law presents an ideal situation which can only be reached gradually. While the results are positive, there are problems, and how serious those are, remains to be seen. In the very least, financing is problematic: in 1994-95 only 22% of needed funds were provided by the state budget, and the Ministry for Nationalities seriously appealed to local administrators to take the programme into consideration in their respective budgets (Spravka o natsional´noi situatsii v Respublike Komi). The same trend seems to continue, judging by the complaints.

In detail, the efforts that aim at maintaining the Komi language and reversing language shift are the following:

1) According to Art. 20 of the Law on official languages, in principle, both languages are to be studied in the comprehensive schools. Further, pupils and students may choose the language of their entrance and final exams and scientific work (Art. 21). However, it remains for each district to decide how to put this into practice, and in reality, the economic situation dictates its own

laws: the hours dedicated to indigenous language are minimal, there are difficulties in producing text-books etc. There are schools, where there are 3-6 hours per week dedicated to teaching the Komi language, and kindergartens with teaching of songs. In 1995/96 some 18% of pupils studied Komi, which certainly shows a progress, compared to 10,5% in 1990/91. There are three schools with an extended curriculum of Komi. Courses for adults are being arranged. There are plans to teach all school subjects in the first four classes in Komi (Spravka o natsional´noi situatsii v Respublike Komi). This has not yet been realized.

2) For the preparation of teachers of Komi, there are new courses at the Pedagogical Institute and the University. While some Komi functionaries claim that there is no need for more teachers, a senior presidential advisor, G. Butyreva (herself a Komi), promises that by the year 2007, there will be teaching in Komi in all schools.

3) At the University of Syktyvkar there is a new Finno-Ugrian faculty and a department of regional studies, where Komi is taught to non-natives beginning a career in the republican administration. A 4 year curriculum leads to a BA degree. There are, however, problems with accreditation. The position of the department is precarious, and it continually gets closed down and re-opened again. According to the Ministry, in all faculties, as well as in schools and professional colleges, a course on the history, literature and culture of Komi is being taught in Russian.

4) According to the law, there are to be newspapers and journals, television and radio transmissions in both languages (Art. 22). In radio and TV, Komi transmissions account for ca. 30% of the programming time of the local news station Komi gor. It does not, however, produce many hours per day, and it competes against several Russian channels. In June 1998, Komi TV

424

programmes amounted to less than 2 hours a week, and the radio programmes are concentrated between 6 and 7 o´clock a.m. and p.m. There is one all-republican Komi newspaper (Komi mu, 3 issues a week, edition dropped from 15,000 to 9,000 in September 1998), a children's weekly journal (Jölöga, edition ca. 2,000), and a children's monthly journal (Bi kin´, edition 5,800 dropped to 1,300). The edition of the old-established literary journal Voivyv kodzuv has fallen as well (1,440 in September 1998, cf. 7,000 in the 70's). There are three local newspapers that carry articles in Komi. A radical opposition party publishes infrequently its own Komi-Russian newspaper Doriam as´nymös. A new bilingual cultural journal Art/Lad has been published since 1997, and three numbers exist so far. As to books, the Komi Publishing House has been privatised, which means strict economising. In 1998, so far only one book has been published in Komi.

5) According to Art. 23 and 25, names of places, streets, official institutions, stamps and letterheads should be in both languages: the results are becoming visible now.

6) According to Art. 18, there should be a list of such administrative positions for which the knowledge of both languages is required. So far, such a list does not exist. State officials of ministries do go to language courses of 7-8 months to learn Komi (Art. 13 names 'a minimum knowledge').

7) Komi should be able to get service in Komi not only in state administration, but also in the shops (Art. 4). The feasibility of this is unknown - no one is trying. At least, some shops have Komi names.

8) According to the Programme, language planning should be carried out. The early normative decisions of the 20's have been forgotten, and variative language use causes problems for school teachers. At the Ministry for Nationalities, there works a terminological-orthographic commission, with the

task of creating Komi neologisms for Russian terms which were used without translation in Komi texts up to now. The results are published in Bulletins, and a booklet of 900 new words came out in 1997. The new words are implemented on the radio and in TV programmes (Tsypanov 1997). Most journalists, however, write as they used to, though some use a Komi word adding in brackets the Russian equivalent.

Among new positive developments may be mentioned the following:

- while the national theatre in the capital tends towards Russian classics, there is a new folklore theatre showing plays in Komi since 1992;
- there are Komi cultural festivals, which were planned for every second year, but now take place every fourth year, while Russian cultural festivals have been started;
- seminars and scientific congresses dealing with Komi are organised at a fast rate;
- a lot on Komi history, culture etc. is being published in Russian, including reprints of earlier research. A new encyclopaedia of the Komi language has just appeared;
- courses and scholarships to study Komi are arranged for students coming from the Komi diaspora and the Komi-Permiak area.

### 5.1. The main problems

As pointed out by Lenore A. Grenoble & Lindsay J. Whaley (1998), there indeed seem to be three crucial issues in minority language maintenance that are more important than any micro-variables:

- economics: the key motivation for the speakers is economic and social advancement on one hand, and on the other, money dictates the availability of published materials, schools, teachers, massmedia, internet etc.;

- access: as being related to economics, the most important is the access to indigenous language and culture in the media;
- motivation, personal and community attitudes.

While the economic and access factors were treated above, the attitudes of the Komi themselves are of extreme importance. There are many individuals in the Komi intelligentsia who doubt the importance of maintaining their native tongue and see it as a obstacle, if not in the face of Russian, then at least as a waste of time and energy that could be spent in a more profitable way.

While the negative attitudes towards the 'aborigines' during the Soviet era were fostered by the prestige of Great Russian Culture and contributed to their assimilation, we now might expect a change due to the recent ethnic revival. The development seems to be slow. For instance, in a survey conducted in 1995, only 43% of Komi and 16% of non-Komi felt that the teaching of Komi should be obligatory (Muller 1998:612-614). In 1993 in a sociological survey, rural Komi were asked why people should retain their indigenous language. Almost all felt that this was necessary, as it was the mother tongue, but at the same time they thought that one should show understanding towards other nationalities - only one third felt that others should know Komi. One out of ten thought that this should be the result of the new law on the official languages. The researchers, themselves Russians, praised the 'positive neighbourly attitude' of the Komi, but did note that it also reflected the low esteem that they have about themselves and their culture. For instance, they felt embarrassed to use their language in public. The conclusion of the researchers was that the functional sphere of the language was diminishing (Denisenko, Kalianova 1994).

## 5.2. Other factors

In the above, I have taken into account the factors affecting the maintenance and loss of minority languages presented by John Edwards, as cited in Grenoble & Whaley (1998:25-26). The following factors therein have a lesser impact on the situation in Komi:

- Gender does not seem to be a significant factor any more.Still, there are slightly more monolingual Komi women than men - bilingualism reaches over 90% (Lallukka 1997:229).
- Religion: The Russian Orthodox church functions in Russian and Church Slavonic only. However, there is an Evangelical Western religious organisation that has undertaken the (re-)translation of the Bible and is spreading it among the Komi - as well as to other minorities in Russia. The results remain unknown.
- Literacy: In Komi in the 70's, the teaching of Russian was the most intensive among the autonomous republics of the USSR. A survey in 1993 of the rural Komi population, which has retained the language best, showed that as the basic language of communication Komi is used only by the oldest generation. 64,6% could write Komi without trouble, 19,4% with difficulty. 8% could not read Komi. Among those who could read and write Komi, only 15,3% read it regularly, and wrote it even less. Still, Komi literature is read, but in Russian translation. (Denisenko, Kalianova 1994).

Purism is a problematic factor. Presumably a large part of Komi consider themselves as semi-speakers, and feel shy of using the language in official situations, as it would draw criticism from normativists, self-appointed and official. A separate problem with purism is regional linguistic distinctions. The

Izhma dialect differs from the literary language, not so much on the grammatical and phonological level as in the lexicon. Due to a Komi-Nenets-Russian population mixture, its speakers complain that they cannot understand texts produced by the centre. One of the results is that they publish their newspaper partly in Russian, partly in their own dialect rather than in literary Komi.

## 6. Conclusion

After 1993, authoritarianism became popular in Russia, and at present, nationalism is strong among the Russians as well. In 1996, the republic began the realisation of the programme for the maintenance and development of the Russian language, representing the other official language. Presumably, this will require much financing and effort as well.

Among the Komi, there are voices saying that the financing seems to be aimed at raising the official 'face' of the republic rather than at grass-level work. No doubt, much of this is due to the newness of the idea of official bilingualism. The mental climate shows not exactly anti-Komi feeling, but, as some complaints go, rather passivity or lip-service to the realisation of the language programme (Tsypanov, ms.).

It may seem astonishing that a parliament with a Russian majority has accepted a law that makes Komi equal with Russian in principle. The motivation is obviously political. A consequence of the raised status of the Komi language is that the state has effectively contained the growth and spread of Komi nationalism (Fryer 1998). In all likelihood, Komi will probably not function as a full-fledged official language, but may remain diglossic, with an extension of its functional sphere to administrative use in some areas. No one is obliged to learn Komi (Art. 18 of the Language Law), while Russian will continue to be absolutely necessary - the percentage of Russians learning another language, including Komi or that of any of the other nationalities, has been for decades ca. 1%. No doubt the decisive factor for language shift reversal will be the urbanised bilingual Komi, who up to now have not deemed it necessary to speak to their children in Komi.

*University of Tampere*
*Filologia II*
*PL 607*
*33101 Tampere, Finland*

**References:**

Denisenko V. and T. Kalianova. 1994. Problemy rodnogo iazyka v srede komi sel´skogo naseleniia. Orientir. Obshchestvenno-politicheskii zhurnal Respubliki Komi. Sbornik nauchnykh rabot, vypusk 2. 1-2:48-50.
Doc. 8126 = Endangered Uralic minority cultures. Report to Committee on Culture and Education, Council of Europe, Parliamentary Assembly, by Mrs. Tytti Isohookana-Asunmaa, Finland, 2.6.1998.
Fryer, Paul. 1998 = Lynn, Nicholas J. and Paul Fryer. National-territorial change in the republics of the Russian North. Political Geography Vol. 17, 5:567-88.
Grenoble, Lenore A., and Lindsay J. Whaley. 1998. Toward a typology of language endangerment. Grenoble, Lenore A., and Lindsay J. Whaley (eds.). Endangered languages. Current issues and future prospects, 22-54. Cambridge: Cambridge University Press.
Kiselev, V. 1998. Piatyi s"ezd komi naroda: itogi i resheniia. Art 1:189-190.

Kratkie svedeniia = Kratkie svedeniia ob istorii razvitiia Komi natsional´noi shkoly. n/d.

Lallukka, Seppo. 1990. The East Finnic Minorities in the Soviet Union. An Appraisal of the Erosive Trends. (Annales academiae scientiarum fennicae. Ser. B:252). Helsinki: Suomalainen Tiedeakatemia.

Lallukka, Seppo 1997. Vostochno-finskie narody Rossii. Analiz etnodemograticheskikh protsessov. Sankt-Peterburg: Evropeiskii Dom.

Muller, Barbara. 1998. Sprachenpolitik in der Republik Komi. Osteuropa 6:604-16.

Popov, A. 1990. 'Zyrianizatsia'. (1918-30-e gody). Komi kyv öniia olömyn. Funktsionirovanie komi iazyka v sovremennykh usloviiakh. 82-93. Syktyvkar.Shtrikhi etnopoliticheskogo razvitiia Komi respubliki. Ocherki. Dokumenty. Materialy.

Tom I. 1994. Ed. N.M. Guboglo. Moskva: Rossiiskaia akademiia nauk, Tsentr po izucheniiu mezhnatsional´nykh otnoshenii im. N.I. Miklukho-Maklaia. Spravka o natsional´noi situatsii v Respublike Komi. Ministerstvo po delam natsional´nostei. n/d.

Tsypanov, E. 1997. Iazykovaia politika i komi literaturnyi iazyk. Art 1:198-206.

Tsypanov, E. (ms.) Lingvoetnicheskie protsessy u komi-zyrian v seredine 90-kh godov.

Zherebtsov, I.L., Savelieva E.A., and A.F. Smetanin. 1996. Istoriia Respubliki Komi. Nauchno-populiarnye ocherki. Syktyvkar.

# Phonetic salience and language change: Verb paradigms in Inari Sami

Ida Toivonen

**Abstract:** This paper presents data on Inari Sami verb morphology. Two informants differ from other speakers in that they have merged the second person dual and plural endings in the present tense paradigm. The two speakers use different endings: one of them uses the former dual ending whereas the other uses the former plural ending. The explanation for how the relevant changes which led to the present systems took place will rely crucially on an acquisition-based theory of historical change.

## 1 Introduction

Inari Sami is one of the three Sami languages that are still spoken in Northern Finland. There remain about 400 speakers, who are all bilingual in Inari Sami and Finnish. Most of the speakers do not read or write Inari Sami, and they have not studied it in school in any form: the language has been maintained orally.[1] There are significant dialectal differences between villages where the language is spoken, and there are also interesting idiolectal differences. This is not surprising, since there is no written tradition and no agreed-upon norm. This sociolinguistic setting is very interesting for the linguist who is interested in microdialects and language change. In this paper, we will examine two particular cases of morphological variation. It will be argued that the changes that have led to this variation are best explained if *language acquisition* and *phonetic salience* are taken into account.

## 2 Second person dual and plural

In the speech of most Inari Samis, and also in the Inari Sami dictionaries (Sammallahti and Morottaja 1993, Itkonen et al. 1986), the second person dual ending is *-vettee* and the second person plural ending is *-vetteð* in the present tense verb paradigm. This is seen in examples (1-2).[2]

(1) *Tuoi* *kuálááśtvettee* *onne.*
you.DUAL fish.2DUAL today
'(The two of) you are fishing today.'

(2) *Tiij* *kuálááśtvetteð* *onne.*
you.PLURAL fish.2PLURAL today
'You (all) are fishing today.'

Two of my informants have merged these two endings into one. Interestingly, they have not merged them in the same way. One speaker, who I will call *Speaker A*, uses the ending *-vettee* for both dual and plural, and another speaker, *Speaker B*, uses the ending *-vetteð* for both.[3] Below we see the complete present tense paradigm as elicited from Speaker A, Speaker B, and other informants.[4]

(3) *kuáláśtið* 'to fish'

| | | SPEAKER A | SPEAKER B |
|---|---|---|---|
| Sg 1 | kuálááśtam | kuálááśtam | kuálááśtam |
| 2 | kuálááśtah | kuálááśtah | kuálááśtah |
| 3 | kuálááśt | kuálááśt | kuálááśt |
| Du 1 | kuálááśteen | kuálááśtáán | kuálááśtáán |
| 2 | **kuálááśtvettee** | **kuálááśtvettee** | **kuálááśtvetteð** |
| 3 | kuálááśtava | kuálááśtava | kuálááśtava |
| Pl 1 | kuálááśtep | kuálááśtâp | kuálááśtâp |
| 2 | **kuálááśtvetteð** | **kuálááśtvettee** | **kuálááśtvetteð** |
| 3 | kuálááśteh | kuálááśteh | kuálááśteh |

Note that there are some similarities between Speakers A and B which are not shared with the other speakers. (e.g. *kuálááśtep* ~ *kuálááśtâp*). This is not surprising, because A and B are siblings and have lived together since childhood. It is, however, surprising that the two siblings differ in the second person endings, although their primary linguistic data (PLD) must have been very similar. The fact that A and B are siblings makes this a particularly interesting case, since this means that we can be certain of two things. First, in the PLD, both the ending *-vettee* and the ending *-vetteð* existed, so their parents were most probably speaking the dialect of most speakers of today; that is, they distinguished between second person dual and plural. Second, we cannot attribute the change to the absence of either form in the PLD. This is clear because if something in the PLD would have caused a change, than A and B should have the same ending for dual and plural, but they do not: one has *-vettee* and the other *-vetteð*.

Below are further examples of the second person dual and plural forms of verbs other than *kuáláástið* 'to fish'.

(4) *lávluð* 'to sing'

| | | SPEAKER A | SPEAKER B |
|---|---|---|---|
| Du 2 | lávluvettee | lávluvettee | lávluvetteð |
| Pl 2 | lávluvetteð | lávluvettee | lávluvetteð |

(5) *sárnuð* 'to speak'

| | | SPEAKER A | SPEAKER B |
|---|---|---|---|
| Du 2 | sárnuvettee | sárnuvettee | sárnuvetteð |
| Pl 2 | sárnuvetteð | sárnuvettee | sárnuvetteð |

(6) *tubdâð* 'to know; to feel'

| | | SPEAKER A | SPEAKER B |
|---|---|---|---|
| Du 2 | tubdâvettee | tubdâvettee | tubdâvetteð |
| Pl 2 | tubdâvetteð | tubdâvettee | tubdâvetteð |

Examples (7-9) show the verb *tubdâð* used in a second person dual context. Example (7) is spoken by speaker A.

(7)

| | | | | |
|---|---|---|---|---|
| *Tun* | *já* | *enni* | **tubdâvettee** | *sunnuu.* |
| you.SG | and | mother | know | them.DU.ACC |

'You and your mother know (the two of) them.'

Example (8) is spoken by Speaker B.

(8) *Tun* *já* *enni* **tubdâvetteδ** *sunnuu.*
you.SG and mother know them.DU.ACC
'You and your mother know (the two of) them.'

Example (9) is elicited from a third speaker:

(9) *Tun* *já* *enni* **tubdâvettee** *sunnuu.*
you.SG and mother know them.DU.ACC
'You and mother know (the two of) them.'

Examples (10-12) show the verb *tubdâδ* used in a second person plural context. First speaker A:

(10) *Tiij* **tubdâvettee** *ustevâd.*
you.PL know friend.2.POSS
'You (all) know your friend.'

Example (11) is a quote from Speaker B.

(11) *Tiij* **tubdâvetteδ** *ustevâd.*
you.PL know friend.2.POSS
'You (all) know your friend.'

Finally, example (12) is spoken by a third informant.

(12) *Tiij* **tubdâvetteδ** *ustevâd.*
you.PL know friend.2.POSS
'You (all) know your friend.'

In sentences (7-12) above, it is clear from the context whether the number is dual or plural. We see that Speaker A consistently uses the ending *-vettee* and Speaker B uses the ending *-vetteδ* in both contexts. Speaker A and Speaker B both recognize that other speakers use the ending that they do not use. However, neither of them recognizes a difference in meaning between the two endings. That is, they recognize that both endings are in use, but they do not recognize the dual/plural distinction that other speakers make for second person in the present tense.

## 3 Phonetic Salience

How do we explain the changes that have happened? One speaker has lost the final [ð], while the other speaker has added an [ð] to a morpheme that previously ended in a vowel. In this section, I will discuss two possible explanations for how these changes took place. The first one concerns a general loss of dual/plural distinctions, and I will argue against this possibility. The second explanation refers to acquisition and phonetic salience, and this is the explanation that I will adopt.

### 3.1 Weakening of the dual/plural distinction

When one first encounters the data presented in Section 2, one might hypothesize that the dual/plural distinction in general is disappearing in Inari Sami, at least for some speakers. One might expect that some speakers, at least Speaker A and B, are merging all the dual/plural endings. More specifically, one might expect that for Speaker A, all the dual endings are taking over the plurals, and for Speaker B, one would expect that the plural endings are taking over the duals. Neither of these assumptions is correct. In this section, I will present several facts which indicate that the dual/plural distinction is still alive and well in Inari Sami, for all speakers.

The examples in Section 2 all pertained to *second person* dual and plural in the present tense paradigm. Consider now first and third person.

(13) *tubdâđ* 'to know; to feel'

| | | Speaker A | Speaker B |
|---|---|---|---|
| Du 1 | tobdeen | tobdeen | tobdeen |
| Pl 1 | tubdâp | tubdâp | tubdâp |

| | | Speaker A | Speaker B |
|---|---|---|---|
| Du 3 | tubdâv | tubdâv | tubdâv |
| Pl 3 | tobdeh | tobdeh | tobdeh |

We see in (13) that all of the informants make a distinction between dual and plural in first and third person. If we compare it to the data in Section 2, we might expect Speaker A to use *tobdeen* for both dual and plural, and Speaker B to use *tobdâp* for both dual and plural, but this is not what we find. Parallel to other speakers, Speaker A and B make a dual/plural distinction in first and third persons present tense. It is thus clear that there is no general weakening of the dual/plural distinction in the present tense.

If we consider other tenses and moods, we find no indication that the second person dual/plural distinction is disappearing. Table (14) shows the second person forms of *kuálástiđ* 'to fish' in the past tense.

(14) *kuálástiđ* 'to fish', past tense

| | | SPEAKER A | SPEAKER B |
|---|---|---|---|
| Du 2 | kuáláástáid | kuáláástáid | kuáláástáid |
| Pl 2 | kuáláástijd | kuáláástijd | kuáláástijd |

We see in (14) that the distinction between second person dual and plural is kept in the past tense. What we saw in section 2 is not the only dual/plural merger we find ever in the Inari Sami verb paradigms. There are other mergers in paradigms of specific stems that are due to morpho-phonological changes. However, although there are other sporadic mergers, there is no general tendency to lose the dual/plural distinction.

Only one of my informants, Speaker A, does not always keep a strict distinction between the two, although she does recognize the difference and knows all the relevant forms.[5] She sometimes uses the plural forms in dual contexts. Sentences (15-17) are examples of her speech.

(15) *Mij kyehtis tánssáp mii usteváin.*
we.PL two dance.1.PL our.PL friend.COMITATIVE
'The two of us dance with our friend.'

(16) *Tun jå enni haldijd vuojjâđ*
you.SG and mother wanted.2.PL swim.INFINITIVE
*jävrist.*
lake.LOCATIVE
'You and mother want to swim in the lake.'

(17) *Pärni jå olmai tobdeh mij.*
child and man know.3.PL us.PL
'You and the man know us.'

Although she often replaces dual forms with plurals, Speaker A recognizes this as being 'wrong' or 'sloppy'. She sometimes corrects herself immediately after she has replaced a dual with a plural. When specifically asked, she can always explain which form is the correct dual and which is the correct plural form. Note that Speaker A's 'slopppiness' cannot be appealed to to explain the *-vettee/-vetted* merger of Section 2. Recall that Speaker A had generalized the former *dual* form *-vettee* to cover both dual and plural. This is the opposite of the pattern of (15-17), where the *plural* has taken over the dual. So Speaker A's tendency to blur the distinction between dual and plural cannot be used as an explanation for the fact that she has replaced the second person plural form with the dual.

In sum, we do not find evidence that the data presented in Section 2 are due to the fact that the dual/plural distinction in Inari Sami is unstable or disappearing.[6]

## 3.2 Phonetic Salience

Following scholars like Mark Hale (1997) and John Ohala (1981), I take it to be important to consider language acquisition to account for language change. The current proposal will be based on the assumption that the changes we see in Section 2 are due to reanalysis in the acquisition process. That is, when Speaker A and Speaker B were acquiring their language, they 'misanalyzed' or 'misparsed' some of the input.[7] The end result is two grammars that are not identical to each other, nor are they identical to the input grammar. In the table below I summarize the morphological differences that are the focus here.

| | PLD | Speaker A | Speaker B |
|---|---|---|---|
| 2nd dual | *-vettee* | *-vettee* | *-vetteδ* |
| 2nd plural | *-vetteδ* | *-vettee* | *-vetteδ* |

An */δ/* in final unstressed position is not phonetically salient and is thus difficult to perceive in the flow of normal speech. It is thus easy to understand how the change which resulted in Speaker A's language could happen: Speaker A simply did not perceive the final, unstressed */δ/* in the ending *-vetteδ*, and she posited the ending *-vettee* in her mental lexicon for second person plural as well as second person dual. Later, when realizing that some speakers have the final */δ/*, she assumed it was a dialectal variant of the ending without */δ/*.

The change of Speaker B is less straightforward. If final unstressed */δ/* is difficult to perceive, how come it was *added* in the dual morpheme? The key to understanding this change also relies on the fact that final unstressed */δ/* is difficult to perceive. Speaker B learns the ending *-vetteδ* for second person plural, and then he hears *-vettee* for second person dual. Knowing that final */δ/* is difficult to perceive, he hypercorrects and posits a sound that is not there: he assumes that the speaker has said *-vetteδ* and posits that as the lexical entry for second person plural as well. See Ohala (1981) for extensive argumentation that the listener (the acquirer) is often a source of language change in the fashion outlined above.

## 4 Conclusion

In this paper, I have presented data on Inari Sami verb morphology which reveal variation between speakers. Two of the informants differ from the norm, but they do not differ in the same way. The analysis provided here for how the relevant changes have taken place relies on an acquisition-based view of historical change. The reason why the sound in question has undergone changes is that it is not phonetically salient and therefore difficult to perceive. Since the sound is easily misperceived, it is not surprising that it has been misanalyzed in the acquisition process.

Ida Toivonen
Department of Linguistics
Stanford University
Stanford, CA 94305-2150
toivonen@csli.stanford.edu

## Notes

[1] Recently there have been attempts to establish a normative form of the language. Inari Sami has been taught in several shools in the Lake Inari region and textbooks have been created for that purpose.

[2] In the dictionaries, the second person plural pronoun is *tij*, not *tiij*. In the speech of the informants that I refer to in this paper, it is *tiij*.

[3] Note that Speaker A has not lost final [δ] everywhere. In all the infinitive forms cited in this paper, for example, she still has the [δ] sound.

[4] The Inari Sami letter -á- symbolizes a central, mid or low, unrounded vowel. The letter *á* is a low, back, unrounded vowel.

[5] Second person in the present tense paradigm is obviously an exception to this, since Speaker A does not recognize the dual/plural distinction there(Section 2).

[6] In this section, I have based my argumentation solely on specific data from my informants. It should also be noted that 'weakening of a distinction' is an incoherent notion. Either a grammar has a distinction, or it does not.

[7] Terms like 'misanalyze' have a negative connotation. The child is of course just *analyzing* the input. 'Misanalyze' here simply means posit a grammar different from the parents' grammar.

## References:

Äimä, Frans. 1914. Phonetik und Lautlehre des Inarilappischen. Erster Teil. Helsinki: Druckerei der finnischen Literaturgesellschaft.

Hale, Mark. 1997. Theory and Method in Historical Linguistics. Ms., Concordia University.

Itkonen, Erkki, Raija Bartens and Lea Laitinen. 1986-91. Inarilappisches Wörterbuch, vol.I-IV. Lexica Societatis Fenno-Ugricae 20. Helsinki: Société Finno-Ougrienne.

Ohala, John. 1981. The listener as a source of sound change. In C.S. Masek (ed.), Papers from the Parasession on Langauge and Behavior, 178-203, Chicago, Illinois. Chicago Lingusitic Society.

Sammallahti, Pekka and Matti Morottaja. 1993. Säämi-suomâ sänikirje. Ohcejohka: Girjegiisá Oy.

Sammallahti, Pekka. 1984. New developments in Inari Lappish phonology. In Hajdú, Peter (ed.), Studien zur phonologischen Beschreibung uralischer Sprachen. Bibliotheca Uralica 7.

University of Victoria Phonetic Database, Version 3. 1994. Inarisaami (Lappish). SAA-1. Data collected and provided by the Laboratory of Phonetics and Speech Technology, Estonian Academy of Sciences.

# Linguistic and political aspects of alphabet-making for a threatened language

*Jan Heegård*

## Abstract

This paper describes a situation where a native speaker of a threatened language, Kalashamon, has developed an alphabet for his language in order to make it resistant to a heavily expanding neighbouring language and culture. After a linguistic characterization of the alphabet I will present some of the sociocultural or political circumstances that lie behind the alphabet project. And it will come clear why this specific situation in reality is far from being an ideal situation as concerns the working for the survival of a threatened language as described by e.g. England (Hale et al. 1992).

## 1.Introduction

During our field work[1] among the Kalasha in the Chitral District in North West Pakistan a local school teacher, Mr. Ingineer Khan, showed and explained to us an alphabet for Kalashamon, the language of the Kalasha, that he had developed himself and which was planned to serve as the written medium for Kalashamon. Besides from being a quick and most valuable introduction to the complex sound system of Kalashamon the alphabet project gave us an insight into those linguistic and political factors that are involved in constructing an alphabet for an unwritten language.

The aim of this paper is to give a brief presentation of these two different sorts of factors. As concerns the linguistic aspect I will show which features and characteristics of other alphabets Ingineer Khan has found convenient also to be present in the Kalashamon alphabet. As concerns the political aspect I will show why any effort for the survival of this threatened language is a rather touchy matter.[2]

## 2. The sound system

Before we take a closer look at the alphabet it will be useful to have an idea of what the sound system looks like.

Kalashamon has five distinctive vowel qualities and with the addition of distinctive nasalization and retroflexion, marked with ~ and _, respectively, the vowel system becomes extensive:

(1) /i, ĩ, i̢, ĩ̢/ /u, ũ, u̢, ũ̢/
/e, ẽ, e̢, ẽ̢/ /o, õ, o̢, õ̢/
/a, ã, a̢, ã̢/
Diphtongs: /aj, aw, ow/

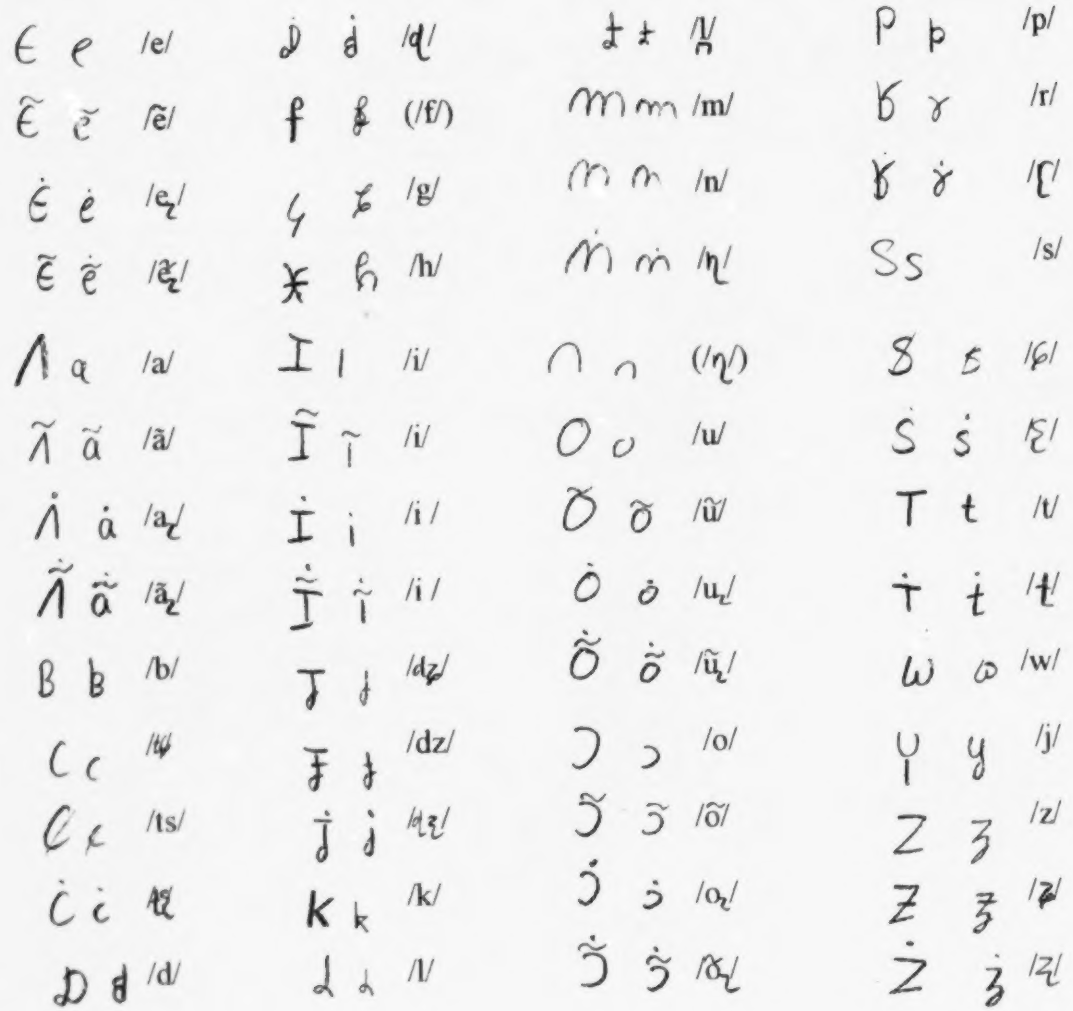
The number of consonant phonemes is large, especially as regards the affricates and the stops with four manners and four places of articulation. The phonemic status of ü and _ can be questioned.

| (2) | **Lab.** | **Dent.** | **Alv-pal.** | **Retro.** | **Pal.** | **Vel.** | **Pharyng.** |
|---|---|---|---|---|---|---|---|
| **Stops** | p | t | | t | | k | |
| | ph | th | | tʰ | | kh | |
| | b | d | | ɖʰ | | g | |
| | bh | dh | | ɖʰ | | gh | |
| **Affricates** | | ts | tɕ | tʂ | | | |
| | | tsh | tɕh | tʂh | | | |
| | | dz | dʑ | ɖʐ | | | |
| | | | dʑh | | | | |
| **Frikatives** | | s | ɕ | ʂ | | | h |
| | | z | ʑ | ʐ | | | |
| **Nasals** | m | n | (ɲ) | ɳ | | (ŋ) | |
| **Liquids** | | | r | | ɽ[3] | | |
| | | l̪ | l | | | | |
| **Glides** | w | | | | j | | |

## 3.Graphic description of the alphabet

Below I have copied Ingineer Khan's alphabet by hand, to be read column by column, and I have given the phonemic value that each letter represents.

444


/e/
/ẽ/
/e˞/
/ẽ˞/
/a/
/ã/
/a˞/
/ã˞/
/b/
/ʧ/
/ts/
/ʈʂ/
/d/
/ɖ/
(/f/)
/g/
/h/
/i/
/ĩ/
/i˞/
/ĩ˞/
/dʒ/
/dz/
/ɖʐ/
/k/
/l/
/m/
/n/
/ɳ/
(/ŋ/)
/u/
/ũ/
/u˞/
/ũ˞/
/o/
/õ/
/o˞/
/õ˞/
/p/
/r/
/ɽ/
/s/
/ʃ/
/ʂ/
/t/
/ʈ/
/w/
/j/
/z/
/ʒ/
/ʐ/

With its specific features and close connection to English, the Kalashamon alphabet appears as a mixture of something unique yet something traditional and familiar. On certain points it is even more coherent graphically than English orthography, i.e. certain letters resemble each other because they represent sounds that are alike. Several features makes the Kalashamon alphabet different from English orthography but only a few will be mentioned here:

a)The use of ~ for nasal vowels and the dot for retroflexion is fully consistent. Notice that the dot above small <i> and small <j> for that reason is missing.

b) <h> alone represents /h/, and together with letters that represent stops and affricates <h> marks aspiration, e.g. <ph> for /ph/, <bh> for /bh/ etc.

c)Among the affricates it is the letter that represents the palatal sound that is the basic letter. Among the sibilants it is letter that represents the dental sound. The reason for this apparent inconsistency might be found in the sounds that the letters represent in English. <s> represents [s], not the palatal [ʃ], which is written <sh>. But because Ingineer Khan consistently uses <h> for aspiration he is left with <s> for the representation of a dental sound and with a modified <s> for /ç /. <c> in English orthography represents in some positions a palatal consonant, and so in Kalashamon. And therefore the dental affricate [ts], which English does not have, is a modified version of <c>.

## 4. The alphabet as a representation of the sound system

Generally, the alphabet follows the principle of one grapheme for one phoneme. If we ignore the digraphic representation of the aspirated sounds there are only few deviations of which two will be mentioned here. (Cf. Heegaard 1996 for others.)

### 4.1 Representation of regular phonetic assimilation

Some words in Ingineer Khan's way of writing show use of both a morphophonemic and a surface phonetic principle, as shown below (LOC = locative case).

| (3) | < > | [ ] | / / |
|---|---|---|---|
| 'branch' | çoŋg | çoŋ | ʻçong |
| 'pear' | toŋg | ʻtoŋ | ʻtong |
| 'hat' | kaşoŋg | kaʻşoŋ | kaʻşong |
| 'horn' | | ʻşiŋ | ʻşing |
| 'horn-LOC' | | ʻşiŋguna | |
| 'cloud' | | ʻmeɲ | ʻmentç |
| 'cloud-LOC' | | ʻmeɲtçuna | |

According to a rule of assimilation a nasal consonant changes its place of articulation according to a succeeding stop or affricate. Another rule applies in word final position where no consonant groups are allowed and therefore the last consonant or consonants are dropped in the pronunciation. This give us pronunciations with a velar nasal as in the words in (3) which have an underlying phonemic structure with a final velar stop. This final stop is realized by suffixation, as seen in the words for 'horn' and 'cloud' in the locative cases. Therefore the use of the letter <ŋ> in the first three words in (3) appears to be a phonetic way of writing. And the use of the letter <g> in the same words is a morphophonemic feature.

### 4.2 Representation of vowel length

A complex feature of Kalashamon phonology is vowel length. In all but one syllable type vowel length is not phonemically distinctive. But in stressed, open, and non-final syllable we find a complex distributional pattern because: a) in one group of words the vowel is always short; b) in another group of words the vowel is always long or near-long; c) in a third group of words the vowel can be either short or long. The difference between a long and a short vowel is exploited in three different ways. (Cf. Heegaard and Mørch 1997, and Heegaard (in press) for a closer description and possible explanations to the distributional pattern of vowel length.) This distribution is exemplified with the near-minimal pairs in (4) below:

(4) a) 'lie on one's side' *ka'patʂa hik* vs. b) 'small feather' *pa:tʂak*
a) 'cold' *tra'mona* vs. b) 'leather bag' *'mo:ni*
a) 'walnut bread' *bi'lili* vs. c) 'ant' *pi'li:lak, pi'lilak*
a) 'him over there' *'a̤la* vs. c) 'pumpkin *'a:luk, 'aluk*

The complicated pattern of vowel length is also reflected in some of Ingineer Khan's word examples to the alphabet. The four first words in (5) below can all be pronounced with a short or a long vowel (classified as 'c' above), but they are written with one vowel letter only. So is the word for 'brother', although it only has a long vowel (classified as 'b' above).

| (5) | < > | [ ] |
|---|---|---|
| 'day' | adu | a:du, adu |
| 'mother' | aja | 'a:ja, 'aja |
| 'grandfather' | wawa | 'wa:a, 'wawa |
| 'grandmother' | awa | 'a:wa, 'awa |
| 'brother' | baja | 'ba:ja |

The words in (6) below further illustrates the complex situation:

| (6) | < > | [ ] |
|---|---|---|
| 'owl' | uuhuk | u'huk |
| 'Prop.' | kanɖ uurik | kaɳ'ɖurik |
| 'lamp' | laʈeɳ | la'ʈe:ɳ, 'laʈeɳ |
| 'poison' | zạr | zạ:r, ar |
| 'lentils' | dal | da:l, dal |
| 'man's sister-in-law' | bahuu | ba'hu:, ba'hu |

In the first two words we have never heard a long vowel but Ingineer Khan writes them as such. In the four other words vowel length varies freely but Ingineer Khan writes them differently: with a short vowel in the words for 'lamp, poison' and 'lentils' and with a long vowel in 'man's sister-in-law'. Altogether the words in (5)-(6) show that Ingineer Khan is uncertain about how to represent vowel length orthographically, a reflection of the complex distributinal pattern of vowel length.

### 4.3 Remarks on the uncertainties

It is worth noticing that these uncertainties and deviations from the otherwise strict phonemic principle used by Ingineer Khan concern the same parts of the sound system that according to Heegaard and Mørch (1997) are most problematic. It should, however, in that respect be taken into consideration that Ingineer Khan's primary inspiration, Latin alphabet/English orthography neither differentiates between e.g. oral and nasal vowels, nor unambiguously differentiates between long and short vowels.

The use of <ŋ> is worth noticing. We saw above that Ingineer Khan has a good intuition of feature analysis, e.g. as shown with the consistent marking of nasal vowels and retroflexion. This well-developed sense of systematizing and feature analysis could have forced him to use a letter for a velar nasal consonant: the stops have four different places of articulation, so must the nasals have. The use of <ŋ> in his alphabet could also be due to influence from Urdu orthography. Corresponding to the Urdu way of writing nasal vowel, with a vowel symbol followed by naam, ں, Ingineer Khan also have examples of <Vŋ>, side by side with the use of <Ṽ> for the representation of a nasal vowel.

An explanation of the uncertainties as regards vowel length could be that Ingineer Khan does not regard vowel length as salient cognitively as the segmental features, the vowel and consonant phonemes, but rather regard it as a sort of prosody that works in accordance with stress which he does not mark. In his book *Writing Systems* from 1987 Geoffrey Sampson expresses similar ideas:

(7) "It is tempting to speculate that contrasts of phonological length (and perhaps suprasegmental contrasts in general) may simply be less noticeable than segmental contrasts to speakers of languages which include them" ... "I suggest that scripts commonly ignore contrasts of length, because they are perceptually less salient than, for instance, vowel quality". (Sampson 1987: 132-133.)

Before one starts speculations about how cognitively salient certain phonological contrasts and prosodic phenomena are for Ingineer Khan one should remember that the orthography is in its experimental stage. It would have been more surprising if Ingineer Khan had not faced such problems that I have mentioned. All in all, I think that Ingineer Khan, from a graphic and linguistic point of view, has created a well thought out and solid alphabet. But however well done, linguistic work on language planning cannot separate itself from political and cultural attitudes, as will become clear below.

## 5.Sociocultural Aspects

It is in light of the cultural and religious threat described in Mørch (this volume) that one should see Ingineer Khan's effort to make Kalashamon a written language. In an interview Mørch and I made with him during our first stay in Chitral, he gave the following arguments for creating an alphabet:

(1) "The Kalash people were isolated before, nobody [: no Kho people] were in the valleys and no language were introduced. [Now] many people have come into the valleys. So as the Kalash culture the language is also on the way to disappear. So if we can write it the next generations can get help. If you will not work on a language then the language will die" (in Heegaard 1996).

It is thus clear that Kalashamon is becoming a written language as a part of a struggle for the survival of the Kalasha culture. It is Ingineer Khan's intention to give the Kalasha the cultural prestige that an orthography can give and that no other language but Khowar has in the multilingual and multicultural district of Chitral. Because the alphabet is based on the Latin or English alphabets and not on a variety of Arabic script the Kalash could show to the rest of the world that they are different from the other (Muslim) tribes in the area. Furthermore, in choosing the Latin alphabet as the model the Kalashamon alphabet can serve as a transitional orthography to e.g. English, and not to Urdu, the other national (and Islam oriented) language of Pakistan.

With his position as a school teacher in one of the few Kalasha schools that have been established recently and with the support of the other school teachers one might think that Ingineer Khan has a good opportunity to fulfill his intentions. But the following citation from a paper given by another prominent Kalasha at an international conference in 1995 shows that not everybody agrees with Ingineer Khan:

(10) "[Some people] think that our language is going to die [and] want to make schools and school books for us in Kalasha. But this language, we speak it all the time. We don't need to be taught what we already know!" (Saifullah Jan 1995ms).

This controversy between two prominent members of the society is even more intriguing because they belong to two different lineages that for decades have been in conflict with each other.

Not only does Ingineer Khan have to take up the dispute with a fellow Kalasha. He also has to convince missionaries from the Summer Institute of Linguistics that they should give up their Urdu based alphabet. An alphabet that they have created in their efforts to make Kalashamon a written language in the first place, and in the second place to translate the Bible and christianize the Kalasha and in that way, according to their conviction, make the Kalasha more capable of resisting the ungoing muslim pressure.

## 6. Concluding remarks

That part of language planning which deal with the implementation of an orthography for the Kalasha is for the reasons mentioned in the foregoing at the moment a rather complicated and sensitive matter. The linguist, or any other well-meaning outsider, cannot avoid taking the position of one of the striding parts if he should work or fight for the its survival. However, with the linguistic work almost solved and finished Ingineer Khan can now start working on the important political aspect.

*The Danish Dictionary*
*The Danish Society of Language and Literature*
*Christians Brygge 1,1*
*DK-1219 Copenhagen K*

## Notes

1. The field work, which has lasted altogether 7 months, could not have been conducted without the financial support of His Royal Highness Crown Prince Frederik's fund, The Danish Research Council of the Humanities, The Faculty of the Humanities of University of Copenhagen, and stud.mag. Per Slomann's Foundation.

2. Cf. Mørch (this volume) for a presentation of the sociolinguistic scenario of the Kalasha.

3. /ø/ is only found in one of the still spoken dialects.

## Literature

Hale, Kenneth et al. 1992. Endangered Languages. Language 68.1.1-43.

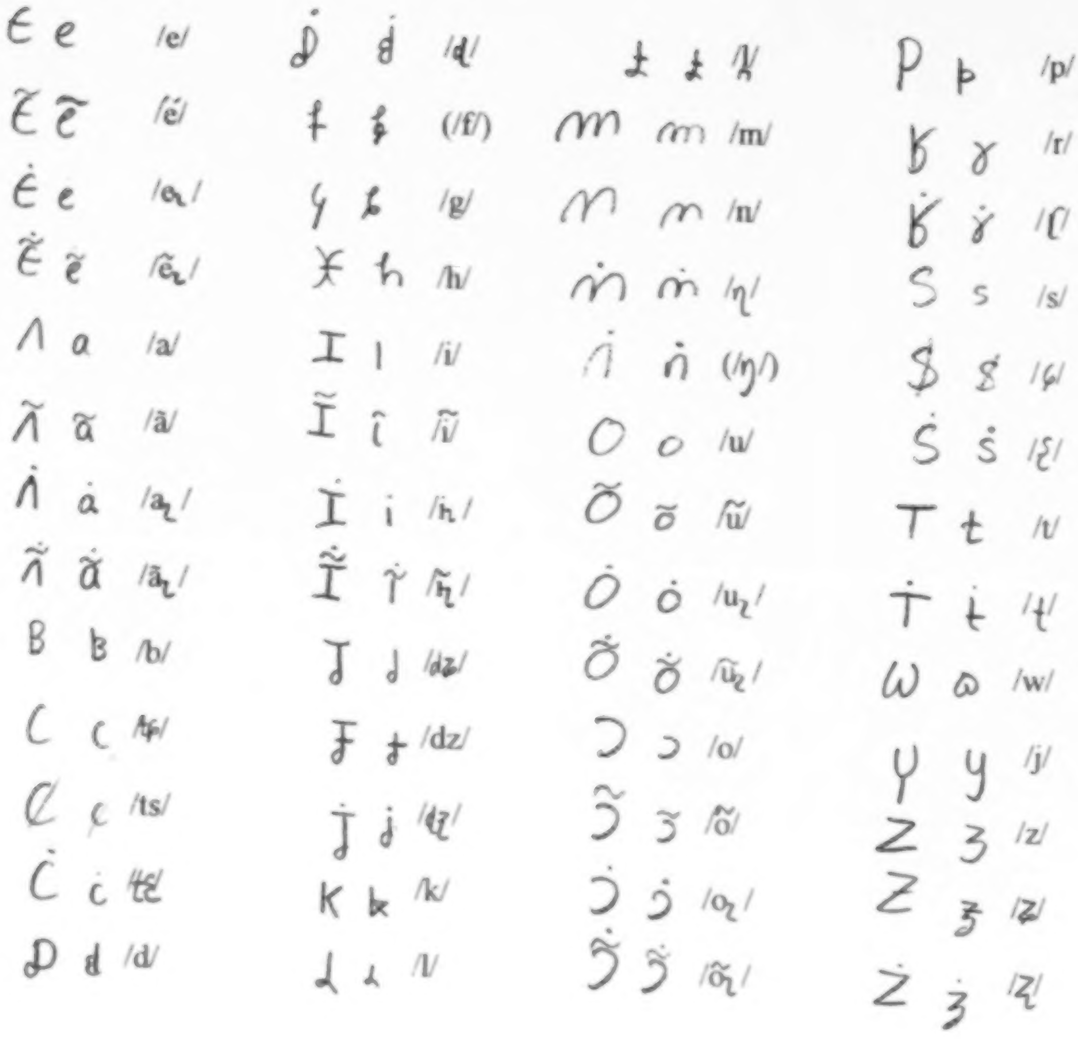
Heegaard, Jan. 1996. Et alfabet bliver til. University of Copenhagen, Department of General and Applied Linguistics. (Term paper.)

Heegaard, Jan. (In press.) Variational patterns in vowel length in Kalashamon. Multilingualism: Basic Issues and Work in Progress, ed. by Jussi Niemi and Terrence Odlin. University of Joensuu.

Heegaard, Jan and Ida E. Mørch. 1997. Retroflekse vokalers oprindelse i kalashamon i historisk og areallingvistisk perspektiv. Variation i sprogbeskrivelsen: vokallængde i kalashamon. To emner i kalashamon-fonologien. University of Copenhagen. MA-thesis.

Jan, Saifullah. 1995ms. (No titel.) To appear in: Proceedings from the 3rd International Hindu Kush Cultural Conference, held in Chitral 1995, ed, by Israr-ud-Din.

Sampson, Geoffrey. 1987. Writing Systems. London: Hutchinson.


Ɛ e /e/
Ɛ̃ ẽ /ẽ/
Ɛ̇ ė /e̢/
Ɛ̃̇ ẽ̇ /ẽ̢/
Λ a /a/
Λ̃ ã /ã/
Λ̇ ȧ /a̢/
Λ̃̇ ã̇ /ã̢/
B b /b/
C c /tɕ/
C c /ts/
Ċ ċ /tʂ/
D d /d/
Ḋ ḋ /ɖ/
F f (/f/)
G g /g/
H h /h/
I ı /i/
Ĩ ĩ /ĩ/
İ i /i̢/
Ĩ̇ ĩ̇ /ĩ̢/
J j /dʑ/
F f /dz/
J̇ j̇ /dʐ/
K k /k/
L l /l/
ł ł /ɬ/
M m /m/
N n /n/
Ṁ ṁ /ɳ/
Ṅ ṅ (/ŋ/)
O o /u/
Õ õ /ũ/
Ȯ ȯ /u̢/
Õ̇ õ̇ /ũ̢/
Ɔ ɔ /o/
Ɔ̃ ɔ̃ /õ/
Ɔ̇ ɔ̇ /o̢/
Ɔ̃̇ ɔ̃̇ /õ̢/
P p /p/
Ɣ ɣ /r/
Ɣ̇ ɣ̇ /ɽ/
S s /s/
$ s /ɕ/
Ṡ ṡ /ʂ/
T t /t/
Ṫ ṫ /ʈ/
Ω ω /w/
Y y /j/
Z ʒ /z/
Z ʒ /ʑ/
Ż ʒ̇ /ʐ/

# WORD OF MOUTH

Communication is power. An idea, passed from person to person, and village to village, can transform the world.

Start with the right idea.

**Linguistics and Language Behavior Abstracts**
offering

- abstracts of scholarly articles and books
- bibliographic entries for subject specific dissertations and book and other media reviews.

LLBA's timely and comprehensive coverage speaks volumes on current ideas in linguistics and language research.

Available in print or electronically through the Internet Database Service (*www.csa.com*). Contact *sales@csa.com* for a trial Internet access or a sample issue.

**LLBa**

***Linguistics and Language Behavior Abstracts***
*Published by Cambridge Scientific Abstracts*
7200 Wisconsin Avenue • Bethesda, Maryland 20814 • USA
+1 301-961-6700 • Fax: +1 301-961-6720 • E-mail: *sales@csa.com*

# Odense Working Papers in Language and Communication

*OWPLC* publishes scholarly articles written by members of the Institute of Language and Communication at the University of Southern Denmark, Main Campus: Odense University and their Danish or foreign colleagues.

Nos 13-19 include:

No. 13 Lene Schøsler & Karen Van Durme:
*The Odense Valency Dictionary*

No. 14 Karl-Heinz Pogner (ed.):
*Writing: Text and interaction*

No. 15 Karin Van Durme (ed.):
*The valency of nouns*

No. 16 Dorthe Bleses & Johannes Wagner (eds.):
*Papers in First Language Acquisition*

No. 17 Dorthe Bleses:
*The role of input, productivity and transparency in Danish children's acquisition of past tense morphology*

No. 18 Teresa Cadierno (ed.):
*L2 Listening comprehension*

No. 19 Carl-Erik Lindberg & Steffen Nordahl Lund (eds.):
*17th Scandinavian Conference of Linguistics, vol I-II*
*20-22 August 1998*

Copies can be obtained from

**Institute of Language and Communication**
**University of Southern Denmark,**
**Main Campus: Odense University**
Campusvej 55
DK-5230 Odense M.
Denmark

General Editor: Steffen Nordahl Lund
(E-mail: snl@language.sdu.dk.)

***Odense Working Papers in Language and Communication*** publishes scholarly articles written by members of the Institute of Language and Communication at SDU - Odense University as well as papers by their Danish or foreign colleagues.

This issue of ***Odense Working Papers in Language and Communication*** contains most of the papers presented at the 17th Scandinavian Conference of Linguistics, which was organised by the Institute of Language and Communication, August 20-22, 1998.

Some 80 linguistics from Scandinavia and other parts of Europe, as well as from the USA and New Zealand, took part in the conference, which adressed the following main topics: The Use of IT in Grammatical Analysis/Parsing; Functional/Cognitive Linguistics; Language in Interaction; Endangered Languages.

UNIVERSITY OF SOUTHERN DENMARK